Analysis of haematopoietic transcription factor networks using TALEs

博 士 論 文

Analysis of haematopoietic transcription factor networks using TALEs

（TALエフェクターを用いた造血転写因子ネットワークの解析）

河田　恵子

平成二十七年度提出

東 北 大 学

SUMMARY

KAWATA, V. K. S. Analysis of haematopoietic transcription factor networks using TALEs. 2014. 84 p. Dissertation (Doctor of Philosophy) - Graduate School of Dentistry, Tohoku University, Sendai, 2014.

Transcription factors (TFs) are key determinants of cell identity and fate, which are thought to act within a highly interconnected TF regulatory network. Numerous TFs including PU.1 are known to play critical roles in developmental and adult haematopoiesis, but how they act within the wider TF network is still poorly understood. Transcription Activator-Like Effectors (TALEs) are a novel class of genetic tool based on the modular DNA binding domains of bacterial plant pathogen *Xanthomonas* TAL proteins, which enable DNA sequence-specific targeting and the manipulation of endogenous gene expression. The work presented in this thesis use engineered TALEs to target the *PU.1-14kb* and *Scl+40kb* transcriptional enhancers, thus providing efficient new tools to perturb expression of these key haematopoietic TFs. It was confirmed the efficiency of these TALEs at the single cell level using high-throughput RT-qPCR which also allowed to assess the consequences of both *PU.1* activation and repression on wider TF networks during developmental haematopoiesis. Finally, combined with comprehensive cellular assays, these experiments uncovered novel for PU.1 during early haematopoietic specification. Therefore, TALEs were established as powerful new tools to study the functionality of transcriptional networks that control developmental processes such as early haematopoiesis.

Key-words: haematopoieisis; transcription activator-like effectors; transcription factor regulatory networks; PU.1

TABLE OF CONTENTS

DEDICATION

I dedicate this work

To my parents

Milton Haruo e Rosalete,

Examples of perseverance, courage and discipline

Mates for life,

That have always guided and supported me,

With unconditional love, dedication and trust.

To you, my eternal gratitude and all my love.

To my sisters and brother

Kelssy Hitomi, Natasha Yuriko and Alan Hiroshi,

By stimulation and affection.

I thank them for the friendly shoulder.

By force that you offer me

making me look up at all times.

My supervisor

Prof. Dr. Hidetoshi Shimauchi,

Whose wisdom gave me a great learning,

Offering me peace of mind to conduct this study.

Always attentive with their dedication,

Helped me open up a wealth of knowledge.

To you, that brilliantly engages in scientific research,

my sincere thanks for further nourish my craving for knowledge.

ACKNOWLEDGEMENTS

My grateful thanks go to Prof. Dr. Hidetoshi Shimauchi not only for allowing me to do this PhD but also for the support he gave over the years that I have spent in his laboratory.

I would also like to express my gratitude to the Japan Society for the Promotion of Science and International Advanced Education and Research Organization of Tohoku University who generously supported me during the course of this work and, research core support grants by the Wellcome Trust to the Cambridge Institute for Medical Research and Wellcome Trust-MRC Cambridge Stem Cell Institute.

Also, I am indebted to Prof. Dr. Berthold Göttgens for his unstinting patience, perseverance and for the molecular knowledge he imparted over the last one year and six months that I have spent in him laboratory. A special thanks to Dr. Sarah Kinston for teaching me how to perform the cellular and molecular experiments, specially for the transfect cells by electroporation and for the many helpful discussions. I would also like to thank Prof. Dr. Bertie Göttgens and Cambridge Institute of Medical Science, University of Cambridge, for allowing me to use his stem cell facility, the help of Dr. Fernando Calero-Nieto, Dr. Nicola Wilson, Dr. Debbie Goode, Dr. Judith Schütte, Dr. Marloes R. Tijssen, Dr. Vasileios Ladopoulos, Dr. Yosuke Tanaka, Dr. Winnie Lau, Jonathan Sive, Victoria Moignard, Dr. David Ruau, Evangelia Diamanti, Dr. Rebbeca Hannah, Felicia Ng, Manuel Sánchez Castillo and Steven Woodhouse has also proved invaluable. Also, I thank Reiner Schulte, Veronika Romashova and Chiara Cosseti for their expertise with cell sorting experiments (from CIMR Flow Cytometry Core),

Cheuk-Ho Tsang (from Prof. Dr. Pentao Liu's laboratory, Wellcome Trust Sanger Institute, Cambridge, UK), Dr. Juan Li and Dr. David Kent (from Prof. Dr. Tony Green's laboratory, Cambridge Institute for Medical Research, Cambridge University, Cambridge, UK) for their helpful suggestions. Finally, special thanks goes to Adam C. Wilkinson whose work and advice was invaluable.

The help of Prof. Dr. Shuntaro Ikawa, Prof. Dr. Yuiko Fujimura and Dr. Yoshihisa Suzuki, our collaborators in Institute of Development, Aging and Cancer of Tohoku University, has also proved invaluable. Finally, special thanks goes to Amin Ruhul whose work and advice was invaluable.

My gratitude to the members of Department of Periodontology and Endodontology, the members of Graduate School of Dentistry Doctoral course, special Prof. Dr. Minoru Wakamori for their professionalism, who conducted the course, and support. I also thank Prof. Dr. Hiroyuki Kumamoto and Prof. Dr. Ozamu Suzuki for having opened the doors of their laboratories and make possible the visualization of all dynamic research conducted in their research fields.

Most of all, I am grateful to my love, Takashi, for his support over the years; his wise words contributed to always put me on a life path even better than I had imagined. Finally, I would like to thank all my family, special Tadashi, Milton Haruo, Rosalete, Kelssy Hitomi, Natasha Yuriko, Alan Hiroshi and all my friends from the four corners of the earth for their help during the happy and difficult periods.

LIST OF ABBREVIATIONS AND ACRONYMS

A260:A280	absorbance at 260 and 280nm
AGM	aorta-gonad-mesonephros
AML	acute myeloid leukaemia
B cell	bursa-derived cells
BFUe	burst forming unit erythroid
Bp	base pair
cDNA	complementary DNA
CFU-G	colony forming unit-granulocyte
CFU-GM	colony forming unit-granulocytes macrophage
CFU-Mix	colony forming unit-mix
ChIP	chromatin immunoprecipitation
Cm	Centimetre
CO_2	Carbon dioxide
Ct	threshold cycle
DAB	3,3′-Diaminobenzidine
DAPI	4',6-diamidino-2-phenylindole
DMSO	Dimethylsulfoxide
DNA	desoxyribonucleic acid
DNAse I	deoxyribonuclease I
Dpc	days post coitus
dNTP	deoxynucleotide triphosphates
Dox	doxycycline
E	embryonic day
EB	embryoid body
EDTA	ethylenediaminetetraacetic acid
EHT	endothelial to haematopoietic transition
ES	embryonic stem
FACS	fluorescence-activated cell sorting
FcR	fragment crystallisable receptor
FCS	foetal calf serum
Ets	E26 transformation-specific
H3K27ac	histone 3, acetylated lysine 27
HSC	haematopoietic stem cell
HSPC	haematopoietic stem/progenitor cell
IU	international units
kb	Kilobase
LIF	leukemia inhibitory factor
MEF	mouse embryonic fibroblast
$MgCl_2$	magnesium chloride
miRNA	microRNA
mg	Milligram

mm	Millimetre
M-MLV	moloney murine leukemia virus
Ml	Millilitre
mRNA	messenger RNA
ncRNA	non-coding RNA
NLS	nuclear localization signal
PCA	principal components analysis
P/S	penicillin/streptomycin
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PFA	paraformaldehyde
PIC	pre-initiation complex
RNA	ribonucleic acid
RNAse	ribonuclease
Rpm	revolutions per minute
rRNA	ribosomal RNA
RT	reverse transcription
RT-qPCR	RT quantitative polymerase chain reaction
rtTA	reverse tetracycline-controlled transactivator
siRNA	small interfering RNA
TALE	transcription activator-like effector
TE	tris-EDTA
TF	transcription factor
tRNA	transfer RNA
U/µl	unit/microlitre
WT	wild type
ηg	Nanogram
ηm	Nanometre
Mg	Microgram
Ml	Microliter
µM	micromolar
° C	degree Celsius

LIST OF FIGURES

LIST OF TABLES

FOREWORD

Unless otherwise indicated, all of the experiments and reagents described in this thesis were performed and prepared by the author. Experimental and procedural contributions made by others include:

OP9 co-culture systems for endothelial assays experiments were performed with the assistance of Dr. Yosuke Tanaka.

Bioinformatic analyses of single-cell gene expression experiments were performed by Adam C. Wilkinson with the assistance of Victoria Moignard.

Partial correlation analysis of single-cell expression experiments were performed by Steven Woodhouse and Jasmin Fisher.

ChiP-seq experiments were completed with Adam C. Wilkinson and analysed by Rebecca Hannah.

Many of the experiments described would not have possible without the generosity of few other investigators. These investigators are individually recognised in the corresponding “Materials and Methods” section.

1 INTRODUCTION

Transcriptions factors (TFs) are key regulators of cell identity and fate. Cell type-specific transcriptional regulation is thought to largely occur by TF binding to distal *cis*-regulatory elements (Heinz et al., 2010). The haematopoietic system provides a well-studied model of mammalian tissue development, in which numerous key TFs have been described [reviewed by Wilkinson and Gottgens (2013)], including Scl (Tal1) and PU.1 (Spi1). The identification of *cis*-regulatory elements that regulate the expression of such TFs has begun to reveal TF circuits that suggest the existence of highly interconnected TF regulatory networks active in the haematopoietic system (Pimanda and Gottgens, 2010; Schutte et al., 2012).

One well-studied example of such haematopoietic *cis*-regulatory element is the *PU.1-14kb* (Rosenbauer et al., 2004; Okuno et al., 2005; Huang et al., 2008; Staber et al., 2013). The *PU.1-14kb* plays a key role in *PU.1* expression in haematopoietic stem/progenitor cells (HSPCs) and mature haematopoietic cell types; its deletion results in an 80% loss of *PU.1* gene expression and acute myeloid leukaemia (AML) in mice (Rosenbauer et al., 2004), while mutation of an (autoregulatory) Ets site within the *PU.1-14kb* causes a 66% reduction in *PU.1* gene expression, which leads to haematopoietic stem cell exhaustion (Staber et al., 2013).

Recent technological advances in microfluidic technology have led to the development of robust protocols for high-throughput quantification of gene expression in single cells (Guo et al., 2010). One of the earliest studies reporting microfluidics-based single-cell

gene expression highlighted the potential for heterogeneity of knockdown efficiency within single cells following siRNA-mediated gene silencing (Toriello et al., 2008). However, the ability to accurately assess gene expression in single cells following conventional perturbations, such as retroviral overexpression or shRNA-mediated knockdown, has been limited because the former commonly yields unphysiologically high expression levels with no means to distinguish between the endogenous and ectopically expressed gene, whereas the latter acts post-transcriptionally and can therefore inhibit protein production without affecting transcript abundance. To realise the full potential of analysing perturbation phenotypes by single-cell gene expression profiling, more physiological means to tune gene expression levels are therefore required.

Transcription activator-like effectors (TALEs) are a novel class of TFs identified in the bacterial plant pathogen *Xanthomonas*, where they are secreted as virulence factors to modulate gene expression of the host plant (Boch and Bonas, 2010). TALEs have a unique modular DNA-binding domain consisting of 33-35 amino acid repeats, each of which binds a single nucleotide with base recognition specificity (Boch et al., 2009). TALEs fused to transcriptional effector domains have been shown to modulate endogenous gene expression (Zhang et al., 2011; Cong et al., 2012; Gao et al., 2013).

The present study use TALEs (fused to transcriptional effector domains) designed to target conserved regions within haematopoietic TF *cis*-regulatory elements as an efficient tool to regulate target gene expression. It was validated TALEs targeting the *PU.1-14kb* element and further assessed the phenotypic effect of modulating the activity

of these enhancers on embryoid body (EB) haematopoiesis. The combination of TALE-mediated endogenous gene expression perturbations with single-cell gene expression studies will be highlighted as a powerful approach to investigate TF regulatory networks.

Osteoimmunology and PU.1

Loss or mutation of lineage regulating transcription factors has yielded insight into the lineage derivation and stages of osteoclast differentiation. As briefly described above, PU.1 is a member of the ETS domain transcription factors that has a key role in regulating the production of B cells, pDC, and all the myelomonocyte-macrophage lineages, including mDC. Mice with targeted deletion of the *PU.1* gene fail to generate monocyte progenitors that express the receptors for GM-CSF, G-CSF, and M-CSF and have severe osteopetrosis due to the complete lack of osteoclasts (Lorenzo et al., 2011). This defect is intrinsic to the osteoclast progenitor as bone marrow transplantation reverses the osteopetrosis in PU.1-deficient mice. *PU.1* expression as the cells differentiated from monocyte to osteoclast, similar to what has been reported for DCs. PU.1 has been demonstrated to interact with the microphthalmia transcription factor (MITF) to regulate *TRAP* (tartrate resistant acid phosphatase) gene expression. Since mice deficient in either M-CSF or its receptor Csf1r are born osteopetrotic but have an age-related recovery of osteoclast production, due largely to the actions of other growth factors, there must be a cell-autonomous function of PU.1 in

the generation of the monocyte-macrophage lineage independent of its role in regulating expression of *Csf1r*. In support of this, *Csf1r* expression (by transduction) cannot rescue macrophage differentiation in PU.1-deficient cells, indicating that, in the absence of Spi1/PU.1, Csf1r signalling is not sufficient to drive macrophage differentiation (Lorenzo et al., 2011).

1.7 Aims of this study

Gene expression is controlled by numerous TFs that bind to *cis*-regulatory regions of their target genes. In order to understand how TFs PU.1 interact to form wider transcriptional networks underlying blood cell development, the following four aims were pursued during the course of the PhD project:

I. Determining the targets and specificity of haematopoietic regulatory element *PU.1-14kb* using TALE;

II. Blocking specific TF binding motifs at this region, to dissect its role enhancer activity;

III. Assess gene function in adult haematopoiesis to test the ability of TALEs to modulate gene expression;

IV. and, perturb TF networks in forward programming experiments.

2. MATERIALS AND METHODS

2.1 Cell culture, sample preparation and classification

2.1.1 Mammalian Cell Line

All cells were grown in controlled conditions using aseptic techniques and good laboratory practice. Cells were maintained at 37 ˚C in a humidified atmosphere containing 5% CO_2 and handled in a class II tissue culture hood. Cells were grown in media as per the advice of the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ) or European Collection of Cell Cultures (ECACC) and maintained at the suggested density. All medium and other reagents were obtained from Sigma Aldrich (Sigma Aldrich Inc, Gillingham, Dorset) unless otherwise stated.

Cell line	Description	Medium
Ainv18 ES	Embryonic Stem Cell	KO-DMEM (Invitrogen), 15% FCS (Gibco), L-Glutamine (Sigma), 100 IU/ml Pen/Strep (Sigma), 2-mecaptoethanol (Invitrogen), recombinated mouse LIF (ORF Genetics), 2.0 x 10^4 inactivated MEFs per cm^2
416b	Mouse Primitive Myeloid Leukaemia cell line	RMPI, 10% FCS, 50 IU/ml Pen/Strep (Moignard et al., 2013)
K562	Human Myeloblastic Leukaemia Cell line	DMEM, 10% FCS, 50 IU/ml Pen/Strep (Knezevic et al., 2011)
OP9	Murine Stromal cell line	α-MEM, 20% FCS, 50 IU/ml Pen/Strep

Table 2.1| Cell line culture conditions.

2.1.2 Estimation of Cell Number

Cell number was determined using CASY impedance counter (Hoffmann-La Roche Ltd, Basel, Switzerland) or the haemocytometer method of counting cells in

solution. This second method was also used to assay live cells by dye exclusion using Trypan Blue (Sigma-Aldrich).

2.1.5 ES cell differentiation

ES cells were differentiated essentially as described in (Sroczynska et al., 2009). At day 4 of differentiation, TALE expression was induced by addition of 0.5 ηg/μl doxycycline and media was refreshed at day 5. EB cultures for single cell gene expression analysis were formed from a mix of WT Ainv18 ES ($mCherry^-$) and targeted Ainv18 ES ($mCherry^+$) cell lines that were passaged once before differentiation as a $^{50}/_{50}$ mixture.

2.1.6 Flow Cytometry

Five-color flow cytometric immunophenotyping was performed on cell line using a 5 laser LSR Fortessa (BD Biosciences). Dissociated EB cells were FcR blocked by incubation with anti-CD16/32 (BD) for 10 minutes at 4 ˚C, then stained with cocktails of monoclonal antibodies conjugated to allophycocyanin (APC), phycoerythrin-cyanine dye (PEcy7), phycoerythrin (PE) or mCherry and directed against the following: CD41-PE-cy7 (Biolegend), CD41-APC (eBioscience), CD45-APC (Biolegend), Flk1-APC (BD), Flk1-PE (BD), VEcad-PE-Cy7 (Biolegend), cKit-APC (BD). DAPI was used as a cell viability stain. Annexin V-APC (BD) antibody and DAPI was used to assess cell apoptosis according to manufacturers' instructions. Results were acquired for 10,000 cells per tube and analyzed using FlowJo software (version 9, Tomy Digital Biology, Japan).

2.2. Cell Biology

2.2.1 Stable transfection

Aliquots of 1.0 x 10^7 K562 or 416b cells were co-transfected with 6 µg of PB-TRP-TALE, 2µg PB-CAT-rtTA and 2 µg of transposase (pl623) by electroporation (BioRad, Gene Pulser X-Cell™ system) The cells were harvested and resuspended in the PBS at 2.5 x 10^7/ml. The plasmids were mixed with 180 µl cell suspension and transfected in 4 mm electroporation cuvettes using a pulse of 220 V at a capacitance of 900 µF and, cells were divided between four 10cm dishes. Cells were subjected to antibiotic selection at 24 hours post electroporation at a dose determined by kill-curve and were assayed once they had re-expanded in number (approximately 10 – 14 days after the addition of antibiotic selection).

For Ainv18 ES cells, cells in log phase growth were co-transfected with 6 µg of PB-TRP-TALE and 2 µg of transposase (pl623) by AMAXA Nucleofector System (Lonza, Slough, UK). 10 µl of plasmids in supplemented Mouse ES cell nucleofector solution were mixed with 90 µl cell suspension and transfected in 4 mm electroporation cuvettes using programme A-024 according to manufacturers instructions. Transfected cells were seeded at single cell density (2.0 x 10^3, 5.0 x 10^3 and 1.0 x 10^4) on gelatin-coated 10cm dishes with 2.0 x 10^4 MEFs per cm^2 and positive clones, based on mCherry expression by fluorescence microscopy and by flow cytometry, were expanded.

2.2.2 OP9 co-culture assays

Flk1$^+$ cells from day 4 EBs (over 95% pure as assessed by flow cytometry) were sorted by MACS (Miltenyi Biotec, Germany) using Flk1-PE and anti-PE microbeads according to manufactures instructions and cultured in 12 well plates containing confluent OP9 stromal cells (Nakano et al., 1994) in MEMα supplemented with 20% FCS at 37 ˚C with 5% CO_2. Mesodermal colonies were allowed to form for 36 hours before doxycycline was added. Haematopoietic cells were not seen before this time point. After 48 hours, cells were fixed in 2% PFA overnight, blocked, stained with purified CD41 antibody (BD), visualize by DAB staining, and mesodermal colonies containing (at least 2) small rounded budding CD41$^+$ haematopoietic cells scored. Specific staining was confirmed using an isotype control antibody. VE-cad$^+$ cells from day 6 EBs were sorted by FACS using a BD Influx and plated into a 12 well plate well containing confluent OP9s, and cultured for 4 days in MEMα supplemented with 10% FCS at 37 ˚C with 5% CO_2. After 4 days, cells were fixed and stained as above using purified CD31 antibody (BD), and endothelial sheet colonies scored.

2.2.3 Endothelial assays

VE-cadherin positive were FACS sorted using a BD Influx and 1000 cells plated into a 12 well plate well containing confluent OP9 cells, and cultured for 4 days in MEM media supplemented with 10% FCS, 2-mecaptoethanol and P/S at 37 ˚C with 5% CO_2. After 4 days, cells were fixed with 2% PFA overnight, stained with purified CD31 antibody, visualized by DAB, and endothelial sheet colonies scored.

2.2.4 Haematopoietic colony forming assays

At day 6, after EBs dissociation using TryLE (Life Technologies), 100,000 cells were plated in triplicate in 1.1ml M3434 Methocult (Stem Cell Technologies) and incubated at 37 ˚C with 5% CO_2. For OP9 co-culture colony forming assays, 100,000 day 6 EB cells were plated in 6 well plate wells on confluent OP9 in MEMα supplemented with 10% FCS at 37 ˚C with 5% CO_2 for 24 hours before media was replaced for M3434 Methocult. Definitive haematopoietic colonies were counted after 10-12 days using the following criteria. Burst forming unit erythroid (BFUe): red coloured erythroid colonies of at least ~30 small cells dispersed within small clusters with tight cell-cell junctions. Colony forming unit-granulocyte (CFU-G): colonies of at least ~50 small round bright cells (often tightly packed with grey centre). Colony forming unit-granulocytes macrophage (CFU-GM): Large colonies of over ~200 cells containing both granulocytes (as described above) and macrophages (large round cell, less bright than granulocytes and often more dispersed). Colony forming unit-mix (CFU-Mix): Large colonies of over ~200 cells, densely packed, including red erythroid cells (similar to those described above) as well as at least two other lineages (usually granulocytes and macrophages described above) or megakaryocytes.

2.3 Molecular Biology

2.3.1 Gene expression analysis

2.3.1.1 RNA Preparation

RNA was prepared from cell lines with 1.0 ml of TriReagent (Sigma, Poole, UK) and frozen at -80 ˚C until needed. For RNA extraction, samples were left to thaw at room temperature and 0.2 ml of chloroform was added per each ml of TriReagent, followed by vigorous shaking. The cells were then incubated at room temperature for 10 minutes, centrifuged at 13,000 rpm for 15 minutes at 4 ˚C and the upper aqueous phase transferred to a new tube. Subsequently, 0.5 ml of propan-2-ol was added to the samples, which were then incubated at room temperature for 10 minutes followed by centrifugation at 13,000 rpm for 10 minutes at 4 ˚C. The supernatant was removed and washed in 1.0 ml of 75% ethanol. The samples were mixed and then centrifuged at 7,500 rpm for 5 minutes at 4 ˚C. Finally, the ethanol was removed, the samples air-dried and the pellet resuspended in 10 − 30 µl of ultra pure sterile water followed by incubation in a heat block set at 55 ˚C for 10 − 15 minutes prior to homogenization the pellet by pipetting up and down several times. 1.5µl RNA was quantified using a Nanodrop and assessment of the absorbance of the samples at 260 ηm and 280 ηm. The purity of the RNA was determined using the ratio of absorbance at 260 ηm to 280 ηm (A260:A280). A ration of close to 1.8 was considered ideal, lower ratios indicated the sample was not fully dissolved and higher ratios indicated protein contamination. RNA

was then DNase treated to eliminate residual genomic DNA using TURBO DNase (Applied Biosystems, Carlsbad, CA, USA) as per the manufacture`s guidance.

2.3.1.2 cDNA Preparation

cDNA was prepared using random hexamers and M-MLV reverse transcriptase reagents kit (Invitrogen) as per manufacturer's guidance.

Briefly, 500 ηg RNA samples were made up to 34.75 µl with RNAse free water in a clean microcentrifuge tube and 10 µl of 10x RT buffer added. Subsequently, 22 µl of 25 mM $MgCl_2$ (final concentration: 5.5 mM), 20 µl dNTPs mix (final concentration: 500 µM per dNTP), 5 µl random hexamers (final concentration: 2.5 µM), 2 µl RNase Inhibitor (final concentration: 0.4 U/µl) and 6.25 µl of MultiScribe Reverse Transcriptase (final concentration 3.125 U/µl) were added and the mixture subjected to thermal cycling with the following conditions. Samples were the stored at -20 ˚C.

Segment	Cycles	Temperature	Time
1	1	25 ˚C	10 minutes
2	1	37 ˚C	60 minutes
3	1	95 ˚C	5 minutes
4	1	4 ˚C	Hold

Table 2.2| Reverse Transcription Thermal Cycling Conditions.

2.3.1.3 qPCR

Quantitative PCRs (qPCR) were undertaken using Brilhant II SYBR Green QPCR Master Mix (Stratagene) following the manufacturer's instructions. Sample were run in Stratagene Mx3000P QPCR System in triplicate using 1 µl of cDNA, 1 µl of forward primer (10 mM), 1 µl of reverse primer (10 mM), 12,5 µl of Brillant SYBR

QPCR master mix and 9.5 µl of ultra pure sterile water. Cycle conditions may be seen on the table below:

Segment	Cycles	Temperature	Time
1	1	95 ˚C	10 minutes
2	40	95 ˚C	15 seconds
		60 ˚C	1 minutes
3	Dissociation curve	95 ˚C	1 minutes
		55 ˚C	30 seconds
		95 ˚C	30 minutes

Table 2.3| Real Time PCR Conditions.

Primers were designed using Primer 3 and Beacon primer-design software and synthesized by Eurofins MWG (Ebersberg, Germany).

Primer name	Sequence
Human ACTB forward	AGAGCTACGAGCTGCCTGAC
Human ACTB reverse	AGCACTGTGTTGGCGTACAG
Human SCL/TAL1 forward	TTCCCTATGTTCACCACCAA
Human SCL/TAL1 reverse	AAGATACGCCGCACAACTTT
Human MAP17 forward	TGCCTATGAGAATGTGCCG
Human MAP17 reverse	TGGACATCCATCCCATGTGC
Human PU.1/SPI-1 forward	CGGCTGGATGTTACAGGCGTG
Human PU.1/SPI-1 reverse	TCGTGCGTTTGGCGTTGG
Human SLC39A13 forward	TTCCCCTAGAGATGGGGACC
Human SLC39A13 reverse	GGCAGCAGATGCAGAAACAC
Human PSMC3 forward	ACAGACGTACTTCCTTCC
Human PSMC3 reverse	CCAATGAACATCTGCACCAG
Human RAPSN forward	GTACGACTCCGCCATGAGCA
Human RAPSN reverse	ATGGCATCCAGAGCCTTGTCC
Human MYBPC3 forward	GCTCTTCCAGACCCATCTCG
Human MYBPC3 reverse	CAGCGGGATGACAGGAAACA
Human MADD forward	TAGTGATCGTAGGGGCCAGG
Human MADD reverse	GCAGGGGAAACTCAGTGTGA
Human STIL forward	ATGCACATAACGTGGATCACG
Human STIL reverse	TCCATGCTCAAATCCACACC
Human CMPK1 forward	TCTCATGAAGCCGCTGGT
Human CMPK1 reverse	TCCTGCAGAAAGGTGTGTGT
Human CYP4X1 forward	TCAGGACACAAGCGTGGAGGTCTA
Human CYP4X1 reverse	TGCATAAGGATCATGGGTGCTGTT
Human CYP4A29-PS forward	CTGCTTTTCAAGGCAGCACA
Human CYP4A29-PS forward	CTGCAAGCAATGCCCAAAGA
Mouse Actb forward	TCCTGGCCTCACTGTCCAC
Mouse Actb reverse	GTCCGCCTAGAAGCACTTGC
Mouse Scl/Tal1 forward	CATGTTCACCAACAACAACCG
Mouse Scl/Tal1 reverse	GGTGTGAGGACCATCAGAAATCTC

Mouse Map17 forward	GTCCTTGTTGCAATCGTCTTC
Mouse Map17 reverse	GAGGAGTATCTGCCATCCATTC
Mouse PU.1/Sfpi-1 forward	AGAGCATACCAACGTCCAATGC
Mouse PU.1/Sfpi-1 reverse	GTGCGGAGAAATCCCAGTAGTG
Mouse Slc39a13 forward	TTGCTGGTCATTCCCCTGGA
Mouse Slc39a13 reverse	GTCCACCTAAGGCAAAGCTGA
Mouse Psmc3 forward	GACCGTGTGGGATGAAGCTG
Mouse Psmc3 reverse	CGCTGGACAATCTCTTCCGTG
Mouse Rapsn forward	ATATCGGGCCATGAGCCAGT
Mouse Rapsn reverse	TCACAACACTCCATGGCACTGC
Mouse Mybpc3 forward	TGAAGGGTCAGTCTCGGTAACC
Mouse Mybpc3 reverse	TCCTGTGGTCGCATCAGAAA
Mouse Madd forward	AAGAAACTGGGCATCCCTCG
Mouse Madd reverse	GAAGGGCACTGGACTTCTCC
Mouse Stil forward	GGTGATGATCAAGAGCCCGA
Mouse Stil reverse	ACCAGGTTCTTTGCTCTGCT
Mouse Cmpk1 forward	TCAGAAGCGCGTTGTATGCT
Mouse Cmpk1 reverse	AAAACGAACACGACCAACGG
Mouse Cyp4x1 forward	CCTGGACATAATAATGAAATGTGCTT
Mouse Cyp4x1 reverse	CTTCACGTAAGACTCATAGGTGCC
Mouse Cyp4a29-ps forward	CAGTGCACCATCTGGACCTC
Mouse Cyp4a29-ps reverse	GATTACGTAATAGTGGTCCCTCAGG

Table 2.4| Human and Mouse Primers.

2.3.2 Single Cell Gene Expression Analysis

Single cell gene expression analysis was undertaken in collaboration with Adam C. Wilkinson.

2.3.2.1 Purification of Progenitor Cells

Cells were pre-incubated with FcR-block for endothelium and haemogenic endothelium stains. Cells were stained with VE-cadherin antibody against mouse antigen to allow separation of the individual population (**Figure 2.1**). A Influx™ Cell Sorter (BD Biosciences) was used for all cell sorting. Chimeric mixture of wild type (WT) Ainv18 control and TALE inducible ES cells, and unstained populations were used as gate-setting controls. Single cells were seeded by an automated cell deposition

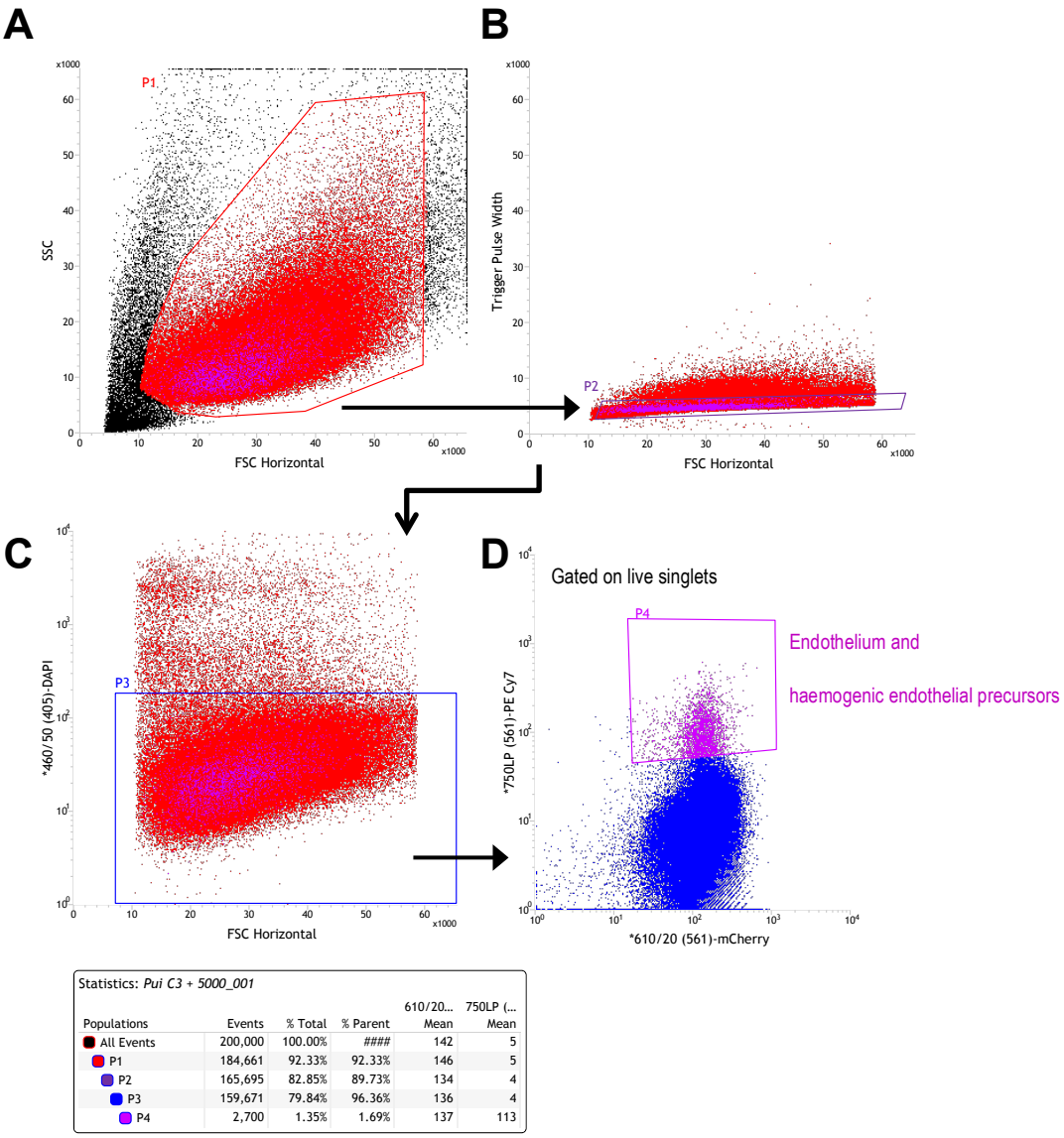
unit directly into to 96 well PCR plates containing lysis buffer, RT/Taq and primers for pre-amplification using a BD Influx (see below).

2.3.2.2 Specific Target Amplification

Single-cell gene expression analysis was undertaken as previously described (Moignard et al., 2013). Single-cell gene expression analysis was performed using 48.48 Dymanic Array integrated fluidics chips (M48, Fluidigm Corporation) on the BioMark HD platform (Fluidigm Corporation), which facilitates the simultaneous analysis of 48 genes in each of 48 samples. Complementary DNA synthesis and specific target amplification (preamplification) of genes of interest were preformed using the CellsDirect One-Step qRT-PCR kit (Invitrogen). Single cells were sorted by FACS directly into individual wells of 96-well plates containing 5 µl CellsDirect 2x reaction mix (Invitrogen), 0.1 µl SUPERase RNAse Inhibitor (Ambion), 2.5 µl 0.2x assay mix, 1.2 µl TE buffer (Invitrogen) and 1.2 µl SuperscriptIII/Platinum Taq (Invitrogen). The 0.2x assay mix contained a pool of 24 TaqMan assays (Applied Biosystems; **see 2.3.2.3**) at a 1:100 dilution of each assay in TE buffer. Reverse transcription and specific target amplification were performed in the sample plates immediately after sorting as follows:

Segment	Cycles	Temperature	Time
1	1	50 ˚C	15 minutes
2	1	95 ˚C	2 minutes
3	22	95 ˚C	15 minutes
		60 ˚C	4 minutes

Table 2.5| Specific Target Amplification PCR Cycling Conditions.



Statistics: *Pui C3 + 5000_001*

Populations	Events	% Total	% Parent	610/20... Mean	750LP (... Mean
All Events	200,000	100.00%	####	142	5
P1	184,661	92.33%	92.33%	146	5
P2	165,695	82.85%	89.73%	134	4
P3	159,671	79.84%	96.36%	136	4
P4	2,700	1.35%	1.69%	137	113

Figure 2.1| Purification of Endothelium and Haematopoietic Endothelium Precursor Cells.

FACS profiles for the sorting of single (P4) endothelium and haematopoietic endothelium precursor cells. Flow cytometry cell events are first gated using FSC-H versus SSC, then singlet events are gated using FSC-H versus trigger probe width. Low Dapi staining is used to gate live cells. mCherry positive and mCherry bright cells are then gated on used Ainv18 cell line in mCherry versus PE-Cy7-VEcadherin staining single stained controls.

cDNA was diluted 1:5 with TE before quantitative PCR (qPCR) on the BioMark HD. cDNA was stored at -20 ˚C before processing on the BioMark HD.

Gene	Assay ID

Csf1r/c-fms – colony stimulating factor 1 receptor	Mm01266652_m1
CD34 – haematopoietic progenitor cell antigen CD34	Mm00519283_m1
CD41 – Integrin, Alpha 2b (Platelet Glycoprotein IIb of IIb/IIIa Complex, Antigen CD41)	Mm00439768_m1
Cdkn2a – cyclin-dependent kinase inhibitor 2A	Mm00494449_m1
Csf2ra – colony stimulating factor 2 receptor, alpha, low-affinity (Granulocyte-Macrophage)	Mm00438331_g1
Eif2b1 – eukaryotic translation initiation factor 2B, subunit 1 alpha	Mm01199614_m1
Epb4.2 – erythrocyte membrane protein band 4.2	Mm00469107_m1
Epcr – protein C receptor, endothelial	Mm00440993_mH
EpoR – erythropoietin receptor	Mm00438760_m1
Erg – V-Ets Avian Erythroblastosis Virus E26 Oncogene Homolog	Mm01214246_m1
Eto2 (Cbfa2t3h) – core-binding factor, runt domain, alpha subunit 2; translocated to, 3	Mm00486780_m1
Ets1 – V-Ets Avian Erythroblastosis Virus E26 Oncogene Homolog 1	Mm01175819_m1
Ets2 – V-Ets Avian Erythroblastosis Virus E26 Oncogene Homolog 2	Mm00468977_m1
Etv2 – Ets variant 2	Mm00468389_m1
Fli1 – Friend leukemia virus integration 1	Mn00484409_m1
Flk1 – Kinase Insert Domain Receptor (A Type III Receptor Tyrosine Kinase)	Mm01222421_m1
GATA1 – GATA binding protein 1	Mn00484678_m1
GATA2 – GATA binding protein 2	Mn00492300_m1
GATA3 – GATA binding protein 3	Mm00484683_m1
Gfi1 – growth factor independent 1 transcription repressor	Mn00515855_m1
Gfi1b – growth factor independent 1B transcription repressor	Mn00492318_m1
Hbb-bH1 – haemoglobin, beta	Mm00756487_mH
Hbb-y – growth factor independent 1B transcription repressor	Mm00433936_g1
hHex – haematopoietically expressed homebox	Mn00433954_m1
Hmbs – Hydroxymethylbilane Synthase	Mm01143545_m1
Hoxb4 – Homeobox B4	Mm00657964_m1
Ikaros – Zinc Finger Protein, Subfamily 1A, 1	Mm01187882_m1
Kit – v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog	Mn00445212_m1
Lmo2 – LIM domain only 2	Mn01281680_m1
Lyl1 – Lymphoblastic leukemia derived sequence 1	Mn00493219_m1
Lyz2 – lysozyme 2	Mm01612741_m1
Meis1 – Meis homebox 1	Mn00487659_m1
Mpl – myeloproliferative leukemia virus oncogene	Mm00440310_m1
Mpo – myeloperoxidase	Mm00447886_m1
Mrpl19 – Mitochondrial ribosomal protein L19	Mn03048937_m1
Myb – V-myb avian myeloblastosis viral oncogene homolog	Mm00501741_m1

Nfe2 – nuclear factor (erythroid-derived 2)	Mn00801891_m1
Notch1 – translocation-associated notch protein TAN-1	Mm00435249_m1
Pecam1 – platelet/endothelial cell adhesion molecule 1	Mm01242584_m1
Pol2ra – polymerase (RNA) II (DNA directed) polypeptide A	Mn00839493_m1
PU.1 (Sfpi1) – spleen focus forming virus (SFFV) proviral integration oncogene spi1	Mn00488142_m1
Runx1 – Runt-related transcription factor 1	Mn01213405_m1
Scl/Tal1 – T-cell acute lymphocytic leukemia 1	Mn00488142_m1
Sox17 – SRY (Sex Determining Region Y)-Box 17	Mm04208182_m1
Sox7 – SRY (Sex Determining Region Y)-Box 7	Mm00776876_m1
Tel – Ets Variant 6	Mm01261325_m1
Ubc – Ubiquitin C	Mn01201237_m1
VE cadherin – Ubiquitin C	Mm00486938_m1

Table 2.6| TaqMan Assays used for Single Cell Gene Expression Analysis.

Housekeeping or marker genes are shaded in grey.

2.3.2.3 qPCR using Fluidigm BioMark HD Platform

For each population, 8 positive controls of 20 cells per well, 24 negative controls (no cell sorted) and 160 single cells were sorted. This corresponds to 4 M48 Dymanic Arrays per population, each containing 2 positive controls, 6 negative controls and 40 single cells. For the qPCR, 3 µl for each TaqMan assay was mixed with 3 µl Gene Expression Assay Loading Reagent (Fluidigm). Then, 2.7 µl of diluted cDNA was mixed with 3 µl 2x TaqMan Universal Mastermix (Applied Biosystems) and 0.3 µl Gene Expression Sample Loading Reagent (Fluidigm). Next, 5 µl of each sample and assay was loaded into individual sample and assay inlets on the M48 Dynamic Array. Samples and assays were then loaded into the reaction chambers of the Dynamic Array using the IFC Controller MX (Fluidigm), and then transferred to the BioMark HD for qPCR.

Segment	Cycles	Temperature	Time
1	1	95 ˚C	10 minutes

2	40	95 ˚C	15 seconds
3	1	60 ˚C	60 seconds

Table 2.7| Quantitative PCR conditions.

2.3.2.6 Hierarchical Clustering and Principal Component

Hierarchical clustering and principal component analysis using the software programming language R (*www.r-project.org*) package library 'ggplots' and its function 'heatmap.2' and 'PC-loading' has been performed by Adam C. Wilkinson.

Hierarchical clustering and principal component analysis were performed only on the data for the 44 transcription factors. Hierarchical clustering was performed using Spearman Rank correlation. Positive and negative correlations between pairs of genes were tested with Spearman Rank correlation, with P-values calculated based on 10,000 permutations. Positive correlations with a Z-score above 12 ($P < 3E-33$) and negative correlations with a Z-score below -4 ($P < 6.09E-05$) were considered significant, with known antagonistic relationships recovered beyond these values. PCA was performed using the pcromp function.

2.3.3 DNA Template for Stable Transfection

2.3.3.1 Design and assembly of Tal Effector Target Sequences

TALE sequences were designed to 20 bp regions within the Scl+40kb and PU.1-14kb elements that were conserved between human and mouse, and were unique within both genomes by BlastN and Blat (Altschul et al., 1990; Kent, 2002). TALEs were assembled and cloned into piggyBac (PB) as described previously by Xuefei Gao

(Prof. Dr. Pentao Liu`s laboratory, The Sanger Institute; Gao et al., 2013), adapted from (Zhang et al., 2011).

3 RESULTS

3.1 Design and validation of TALEs targeting conserved regions within haematopoietic enhancers

The *Scl+40kb* and *PU.1-14kb* element were aligned, and perfectly conserved regions between human and mouse were identified computationally. TALEs were designed to match these regions but nowhere else in the mouse and human genomes so that they could be used in either organism (**Figure 3.1a, 3.2a**). TALE was initially assembled so that they were fused to the VP64 (transcriptional activator) domain (Beerli et al., 1998) and a mCherry fluorescent reporter via a 2A peptide (**Figure 3.1a**). The 2A peptide is proteolytically cleaved upon translation, releasing mCherry as a marker of TALE expression. The TALE constructs were cloned into piggyback transposon-based plasmids (Wang et al., 2008), for efficient stable genomic integration, and under the control of a Tetracycline Responsive Promoter (TRP), to provide inducible expression regulated by rtTA and tetracycline/doxycycline (dox; **Figure 3.1a-c**). Initially, the ability of these TALE-VP64 proteins to modulate target gene expression in both human and mouse systems by expression in K652 (a human erytholeukemia cell line; (Lozzio et al., 1981), and 416B (a mouse myeloid progenitor cell line; Dexter et al., 1979) were validated. In human K562 cells, the TALE-VP64 targeting *Scl+40kb* upregulated *SCL* expression approximately 4-fold but had little effect on *MAP17* expression (**Figure**

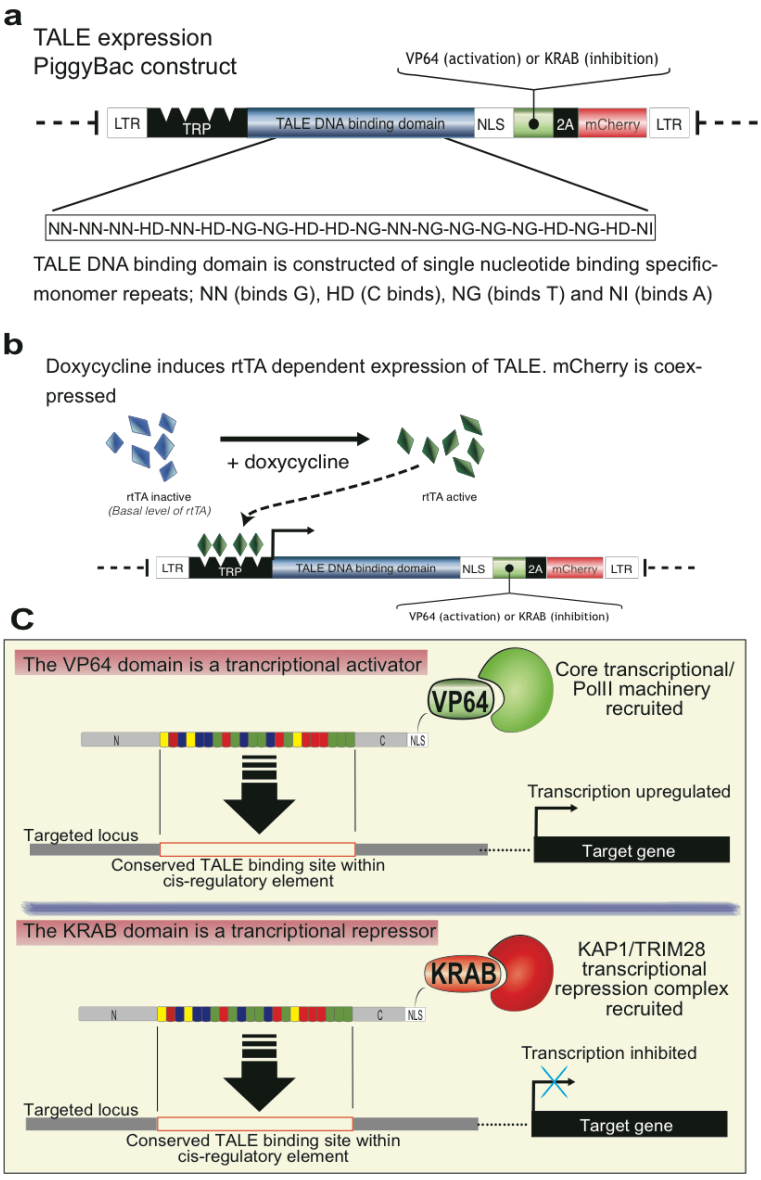
3.2e). By contrast, in 416B cells this TALE-VP64 highly unregulated *Map17* expression ~22-fold but had little effect on *Scl* expression (**Figure 3.2e**). In both the human K562 and mouse 416B cells, expression of the TALE-VP64 targeting the *PU.1-14kb* upregulated *PU.1* expression 3 – 4-fold and *SLC39A13/Slc39a13* expression ~2-fold (**Figure 3.2c**).

Modest (1.5 – 8.5-fold) increases in histone 3 lysine 27 acetylation (H3K27Ac), an epigenetic modification associated active regions of chromatin (Creyghton et al., 2010), were also seen in 416B cells at the promoters of TALE-VP64 target genes, consistent with increased active transcription (**Figure 3.3a, b**). H3K27Ac was also enriched 3.8-fold at the *Scl+40kb* when the TALE-VP64 targeting this enhancer was expressed (**Figure 3.3a**). However, a 50% reduction in H3K27Ac was seen at the *PU.1-14kb* when the TALE-VP64 targeting this enhancer was expressed (**Figure 3.3b**), perhaps due to nucleosome displacement caused by TALE-VP64 and co-factor DNA binding. In mouse embryonic stem (mES) cell, where these enhancers were not active (as determined by H3K27Ac ChIP-seq enrichment; Wilkinson A. C. et al., 2014 unpublished results), and target genes are weakly expressed, TALE-VP64 expression did not induce gene expression upregulation (**Figure 3.3c, d**).

To determine the specificity of these TALEs were further determined expression changes to genes within ~100kb of the target regions (**Figure 3c-f**). Less than 1.7- fold increases in expression were seen in 416B, mES and K562 cells. Reduced expression in some genes (such as *Stil* in 416B cells expressing T-VP64-*Scl+40*) were

identified, perhaps due to transcription factory reallocation (Papantonis and Cook, 2013). TALEs were binding to their regions were additionally confirmed by chromatin immunoprecipitation (ChIP). By ChIP-qPCR, enrichments of 15-17 fold, relative to IgG controls, were seen at target locations (**Figure 3a,b**). To further assess TALE binding specificify genome-wide, the HA-T-VP64-*PU.1-14* sample was sequenced (ChIP-seq) and the 416B control. The number of regions across the entire genome that showed enrichment was comparable between the control IgG and HA-T-VP64-*PU.1-14* samples, underlining the high specificity afforded by TAL-mediated targeting that has also been reported by others (Mali et al., 2013). Importantly, a clear peak at the *PU.1-14kb* element could be identified in the HA-T-VP64-*PU.1-14* sample, but not the control (**Figure 3.4d**). Manual assessment of enrichment at regions containing similar DNA sequences to the TALE-VP64-*PU.1-14* target sequence did not identify strong *off*-target binding, and no either binding events occurred within a 15 Mb window around *PU.1-14kb* (**Figure 3.4d**).

Figure 3.1 | Schematic of TALE Transcriptional Factor Design.



***(a)**, Structure of TALE-expressing piggyBack construct. TALE cDNA consists of TALE sequence followed by a nuclear localisation signal (NLS), the transcriptional effector domain (VP64 or KRAB) and mCherry fluorescent protein, via a 2A (peptide sequence cleaved after translation). TALE cDNA was cloned downstream of a tetracycline-responsive promoter (TRP), and within piggyBack LTRs for stable transposase-mediated genomic integration. The DNA binding domain (DBD) within the TALE sequence consists of twenty monomers; each monomer binds a*

single nucleotide with base specificity, either NN, NI, NG or HD. Sequence of monomer repeats assembled for TALEs targeting PU.1-14kb elements shown below.

***(b)**, Doxycycline activates rtTA to bind and promote expression at the TRP. After protein translation, the 2A peptide is cleaved in vivo to release mCherry, allowing expression to be followed. **c**, the VP64 domain recruits core transcriptional machinery to promote transcription while the KRAB domain recruit transcriptional repressors including NuRD and SETDB1 via KAP1/TRIM28.*

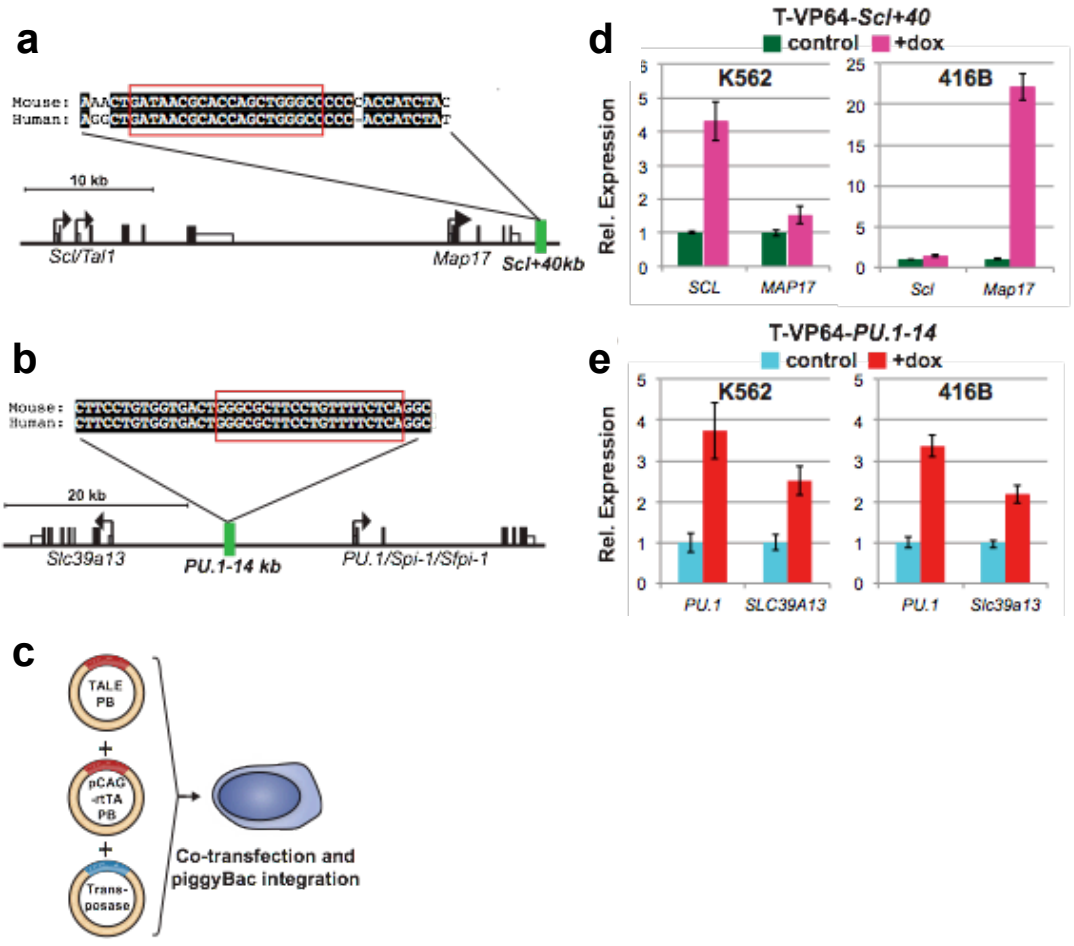


Figure 3.2 | Experimental Validation.

(a) *Schematic of the Scl (encoded by Tal1) genomic locus. Scl+40kb element is highlighted in green, downstream of the neighbouring gene Map17. TALE target site within a conserved (between human and mouse) sequence of the Scl+40kb element highlighted in red.*

(b) *Schematic of the PU.1 (encoded by Spi-1) genomic locus. PU.1-14kb element is highlighted in green. TALE target site within a conserved (between human and mouse) sequence of the PU.1-14kb element highlighted in red.*

(c) *Experimental approach to express TALEs in cell lines. Cells were co-transfected with the TALE expression piggyBack (TALE-PB) from (**Figure 3.1**), a constitutively expression rtTA piggyBack vector (pCAG-rtTA-PB) and a piggyBack transposase, to create inducible TALE expressing cells.*

(d) *Effect of expressing TALE-VP64 targeting the Scl+40kb (T-VP64-Scl+40) in human K562 (left) and mouse 416B (right) on SCL/Scl and MAP17/Map17 gene expression, normalised to ACTB/ActB. T-VP64-Scl+40 expressed for 48 hr by addition of doxycycline (dox) and gene expression in mCherry^{+} cells determined relative to mCherry control cells. Error bars are standard deviation of technical triplicates.*

(e) *Effect of expressing TALE-VP64 targeting the PU.1-14kb (T-VP64-PU.1-14) in human K562 (left) and mouse 426B (right) on PU.1 and SLC39A13/Scl39a13 gene expression, normalised to ACTB/ActB. T-VP64-PU.1-14 expressed for 48 hr by addition of dox and gene expression in mCherry^{+} cells determined relative to mCherry^{-} control cells. Error bars are standard deviation of technical triplicates.*

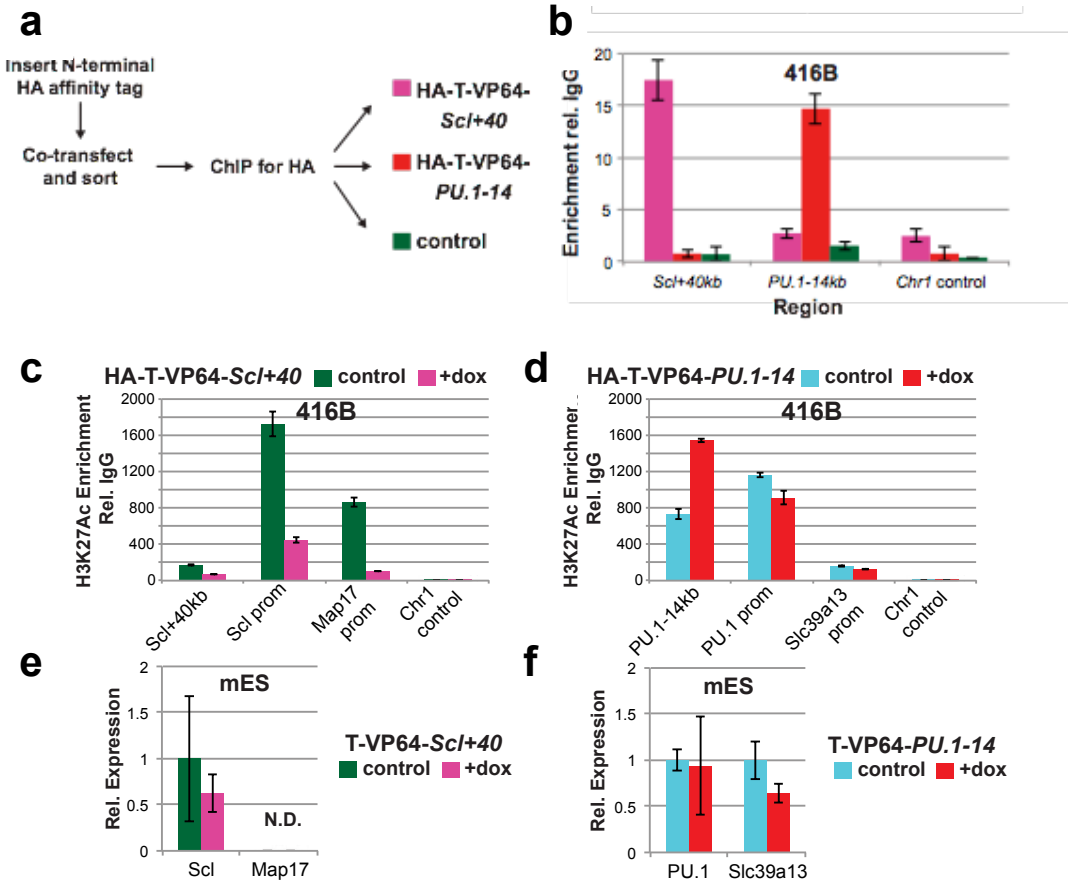


Figure 3.3 | Experimental validation, related to Figure 3.2.

(a) *Schematic of approach taken to ChIP for TALE-VP64 proteins in (H). HA affinity tag was inserted at the N-terminus of the TALE-VP64 protein (HA-TALE-VP64), 416B cells co-transfected as in (D), sorted and ChIP performed 48 hours after dox addition.*

(b) *Chromatin immunoprecipitation (ChIP)-qPCR enrichment of HA-tagged TALE-VP64 (HA-T-VP64) relative to IgG controls in HA-T-VP64-Scl+40 expressing (pink), HA-T-VP64-PU.1-14 expression (red) or untransfected control (green) 416B cells at the Scl+40kb, PU.1-14kb and a control region on chromosome 1 (chr1). Error bars are standard deviation of technical triplicates.*

(c) *Effect of expressing HA-TALE-VP64 targeting the Scl+40kb (HA-T-VP64-Scl+40) for 48 hr in mouse 416B on histone 3 lysine 27 acetylation (H3K27Ac; relative to IgG enrichment) at the*

Scl+40kb, Scl promoter (prom), Map17 prom and a control region on chromosome 1 (chr1 control). HA-T-VP64-Scl+40 expressed for 48 hr by addition of doxycycline (+dox; pink bars). Untransfected 416B used as control (green bars). Error bars are standard deviation of technical triplicates.

***(d)** Effect of expressing HA-TALE-VP64 targeting the PU.1-14kb (HA-T-VP64-PU.1-14) in mouse 416B on histone 3 lysine 27 acetylation (H3K27Ac) ChIP-qPCR enrichment (relative to IgG enrichment) at the PU.1-14kb, PU.1 prom, Map17 prom and Chr1 control region. HA-T-VP64-PU.1-14 expressed for 48 hr by addition of dox (+dox; red bars). Untransfected 416B used as control (blue bars). Error bars are standard deviation of technical triplicates.*

***(e)** Effect of expressing TALE-VP64 targeting the Scl+40kb (T-VP64-Scl+40) in mouse Ainv18 ES cells on Scl and Map17 gene expression, normalised to ActB. T-VP64-Scl+40 expressed for 48 hr by addition of dox and gene expression in +dox cells (green bars) determined relative to –dox control cells (pink bars). Error bars are standard deviation of technical triplicates.*

***(f)** Effect of expressing TALE-VP64 targeting the PU.1-14kb (T-VP64-PU.1-14) in mouse Ainv18 ES cells on PU.1 and Slc39a13 gene expression, normalised to ActB. T-VP64-PU.1-14 expressed for 48 hr by addition of dox and gene expression in +dox cells (blue bars) determined relative to –dox control cells (red bars). Error bars are standard deviation of technical triplicates.*

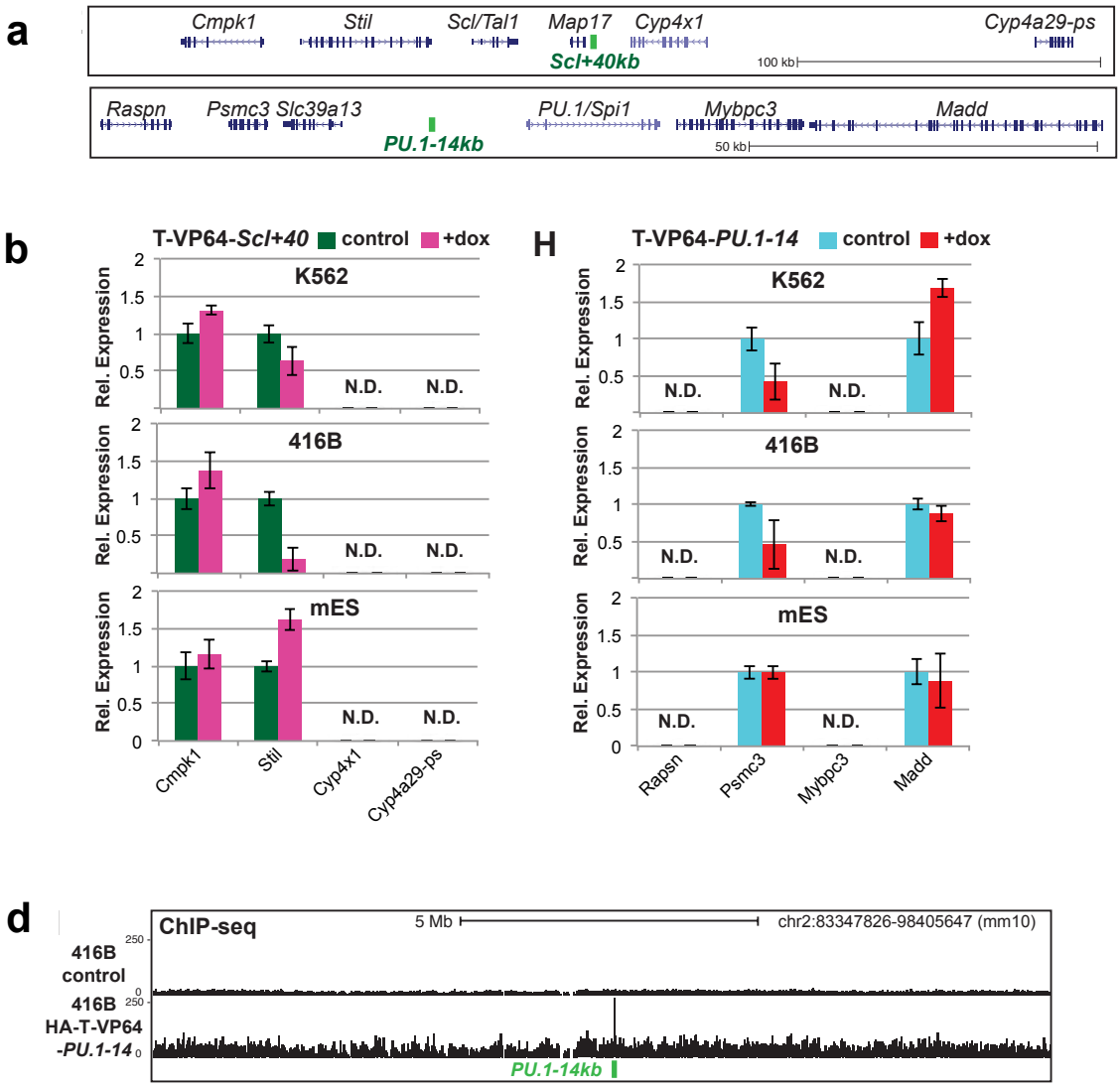


Figure 3.4 | Experimental Validation, related to **Figure 3.2**.

(a) *UCSC Genome Browser screen shot of RefSeq annotated genes within a ~150kb genomic window surrounding the PU.1-14kb enhancer (highlighted in green).*

(b) *Effect of expressing T-VP64-PU.1-14 in human K562 (top), mouse 416B and mouse Ainv18 ES cells on Rapsn, Psmc3, Mybpc3 and Madd gene expression, normalised to ACTB/ActB. T-VP64-PU.1-14 expressed for 48 hr by addition of dox and gene expression on*

+dox/mCherry$^+$ cells (blue bars) determined relative to –dox/mCherry$^-$ control cells (red bars). Error bars are standard deviation of technical triplicates (not detected: N.D).

***(c)** Peak at the PU.1-14kb element could be identified in the HA-T-VP64-PU.1-14 sample, but not the control.*

Having validated the ability of TALEs targeting the *Scl+40kb* and *PU.1-14kb* elements to modulate endogenous gene expression in haematopoietic cell lines, the ability of TALEs to regulate target gene expression was assessed during development. To do so, the mouse EB differentiation system was used to generate haematopoietic progenitors. EBs are spheroid cell aggregates formed by embryonic stem (ES) cells upon differentiation, in which all three germ layers can form and follow normal developmental trajectories (Keller et al., 1993). EB differentiation has been validated as a useful and tractable *in vitro* model of embryonic haematopoiesis (Keller et al., 1993).

The mouse ES cell line Ainv18 (Kyba et al., 2002), which constitutively expresses rtTA from the *Rosa26* locus, was transfected and expanded stably integrated clones that displayed robust mCherry expression after 24hr post dox treatment (**Figure 3.5b**). The data described below is a representative of multiple clones tested for TALE construct. TALE-containing ES cell lines were differentiated, induced TALE expression by addition of dox at day 4 (just prior to definitive haematopoiesis in this system) and assesse phenotypic effects after a further 48 hours of culture, all relative to a culture without dox treatment (-dox) (**Figure 3.5a**). Initial flow cytometric analysis of the day 6 EBs confirmed pure mCherry$^+$ populations at day 6 in the +dox cultures (**Figure 3.5a, c**). Following this protocol, Scl expression was upregulated by approximately 1.9-fold

in cells induced to express the TALE-VP64 targeting the *Scl+40kb*, while *Map17* expression was upregulated over 3-fold (**Figure 3.6a**). *PU.1* expression was upregulated over 4- fold by the TALE-VP64 targeting the PU.1-14kb with no significant expression change in the *PU.1* flanking gene Slc39a13 (**Figure 3.6a**). ES cells containing a *PU.1-14kb* targeting TALE for *PU.1* repression were additionally generated by swapping the VP64 activation domain for the KRAB repressor domain (TALE-KRAB). Following the same EB differentiation and dox-induction protocol as above, it was observed that PU.1 expression was efficiently repressed by the TALE-KRAB, with expression reduced by over 50% (**Figure 3.6a**). The *Slc39a13/ZIP13* gene upstream of the PU.1-14kb was unaffected by TALE expression suggesting that at least in this developmental context, PU.1-14kb activity is specific to *PU.1* (**Figure 3.6a**).

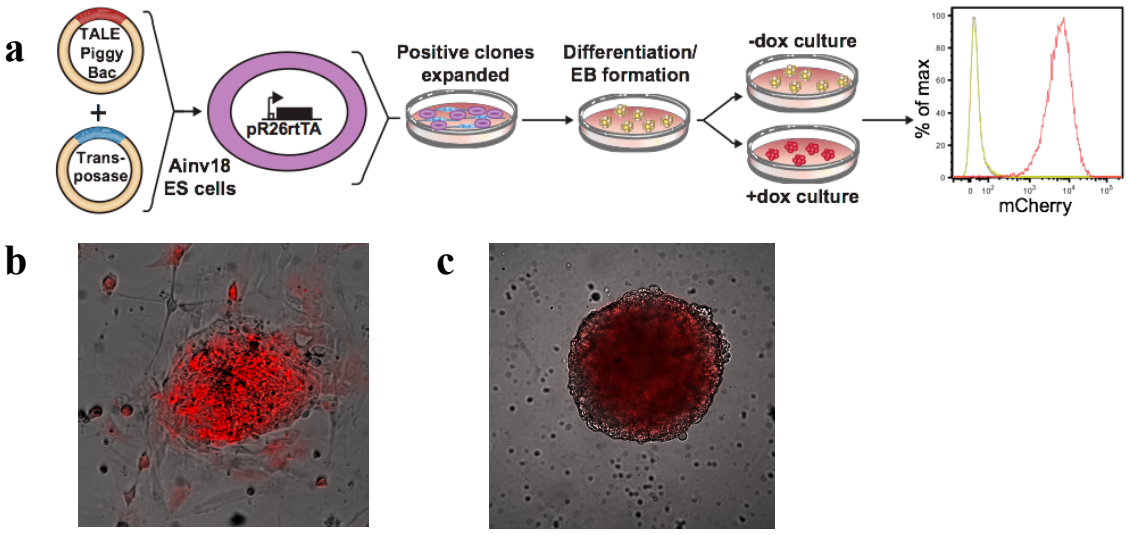


Figure 3.5 | Stable Transfection of ES Cells Using Piggyback Expression Vector.

(a) *Experimental approach using Ainv18 ES cell differentiation to study TALE-mediated gene expression perturbations in haematopoiesis. Mouse Ainv18 ES cells (constitutively expressing rtTA from the Rose26 locus (pR-26-rtTA) were co-transfected with the inducible TALE-PB*

construct and transposase. Stably integrated ES cell clones were expanded and cultured on inactivated MEFs. Targeted ES cells were differentiated into embryoid bodies (EBs), and TALE expression induced at day 4 by addition of dox. At day 6, EBs in +dox culture consistent of a pure mCherry^{+} population, as displayed in the representative histogram (right). Changes in gene expression, colony potential and surface marker phenotype were analysed at day 6 in the +dox expression in day 6 EBs.

(b) *TALE targeted ESC mCherry expression after 24hr post dox treatment.*

(c) *TALE targeted EB mCherry expression after 24hr post dox treatment.*

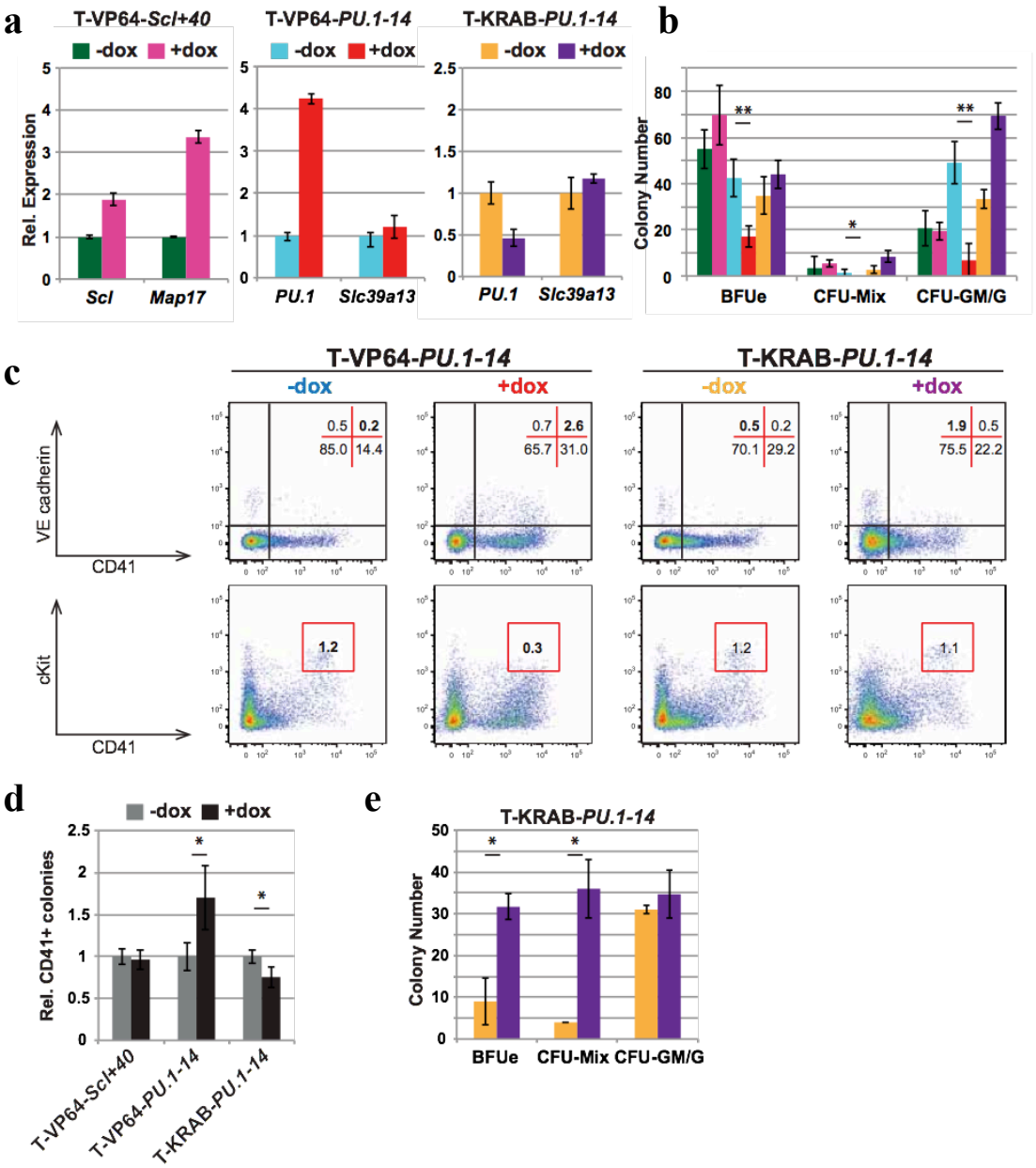


Figure 3.6 | Transient TALE Expression Affects Haematopoietic Cell Fate Decision.

(a) *Gene expression in day 6 EBs of Scl and Map17 after induction of TALE-VP64 targeting the Scl+40kb element (T-VP64-Scl+40; left panel), PU.1 and Scl39a13 after induction of TALE-VP64 targeting the PU.1-14kb element (T-VP64-PU.1-14; middle panel), and PU.1 and Scl39a13 after induction of TALE-KRAB targeting the PU.1-14kb element (T-KRAB-PU.1-14; right panel). TALE expression induced on day 4 (+dox) and displayed relative to uninduced*

(-dox) EBs, normalised to ActB. Error bars are standard error of the mean of three biological replicates.

***(b)** Representative haematopoietic colonies numbers from 1 x 10^5 day 6 EB cells, colour scheme as in (**a**). Colonies grown in methylcellulose supplemented with SCF, IL-3, IL-6 and Epo. See **Figure 3.7a** for images of representative CFU colonies scored. Error bars are standard deviation of technical triplicates. Statistically significant changes ($p < 0.05$ or 0.01) in colony number from three biological replicates determined using the student t test are denoted by * and **, respectively.*

***(c)** Flow cytometry plots of day 6 EB cells showing Flk1 vs. CD41 (top) and VE-cadherin (VEcad) vs. CD41 (bottom). Representative staining patterns for TALE-VP64 (left) and TALE-KRAB (right) targeting PU.1-14kb clones, both uninduced (-dox) and induced (from day 4; +dox). Distribution of cells within quadrant/gates shown as percentages.*

***(d)** Relative number of day 4 EB Flk1+-derived colonies containing CD41+ haematopoietic cells grown on confluent OP9 stromal cells for 84 hours (dox added to +dox wells after 36 hours). See Figure S2G for representative scored colony image. Error bars are standard error of the mean from three biological triplicates. Statistically significant changes ($p < 0.01$) in colony number from three biological replicates determined using the student t test are denoted by *. Grey bars, -dox; black bars, +dox.*

***(e)** Average numbers of haematopoietic colonies from T-KRAB-PU.1-14 1 x 10^5 day 6 EB cells plated onto confluent OP9 stromal cells for 24 hours before CFU assay initiated by addition methylcellulose supplemented with SCF, IL-3, IL-6 and Epo. Error bars are the standard deviation of three biological replicates. Colour scheme as in (a). Statistically significant changes ($p < 0.05$) in colony number from three biological replicates are determined using the student t test denoted by *.*

3.2 Transient Expression of a PU.1 Enhancer Targeting TALE Alters Embryoid Body Haematopoiesis

Next, the phenotypic effect of TALE-mediated gene expression modulation was assessed by haematopoietic colony forming assays using day 6 EB cells (**Figure 3.7a**). PU.1 is known to play a key role in haematopoietic differentiation in the adult bone marrow (Dakic et al., 2007), with PU.1 hypomorphs displaying myeloid differentiation defects leading to AML while PU.1 overexpression causes growth arrest and terminal differentiation (Rosenbauer et al., 2004; Mak et al., 2011). However, comparatively little is known about any possible functions of PU.1 during developmental haematopoiesis. TALE-VP64 mediated *PU.1* upregulation resulted in a significant loss of colony forming ability in day 6 EBs (**Figure 3.6b**). By contrast, TALE-KRAB mediated *PU.1* repression caused a doubling in myeloid (CFU-GM/G) and mixed colony (CFU-Mix) numbers (**Figure 3.6b**), although this was not statistically significant. Colony potential of day 6 EBs were largely unaffected by expressing the TALE-VP64 targeting the *Scl+40kb*, besides a slight (but not significant) increase in BFUe frequency (**Figure 3.6b**). These results confirm that TALE expression in EBs can reveal cellular phenotypes caused by the induced gene expression changes of key regulators such as Scl and PU.1.

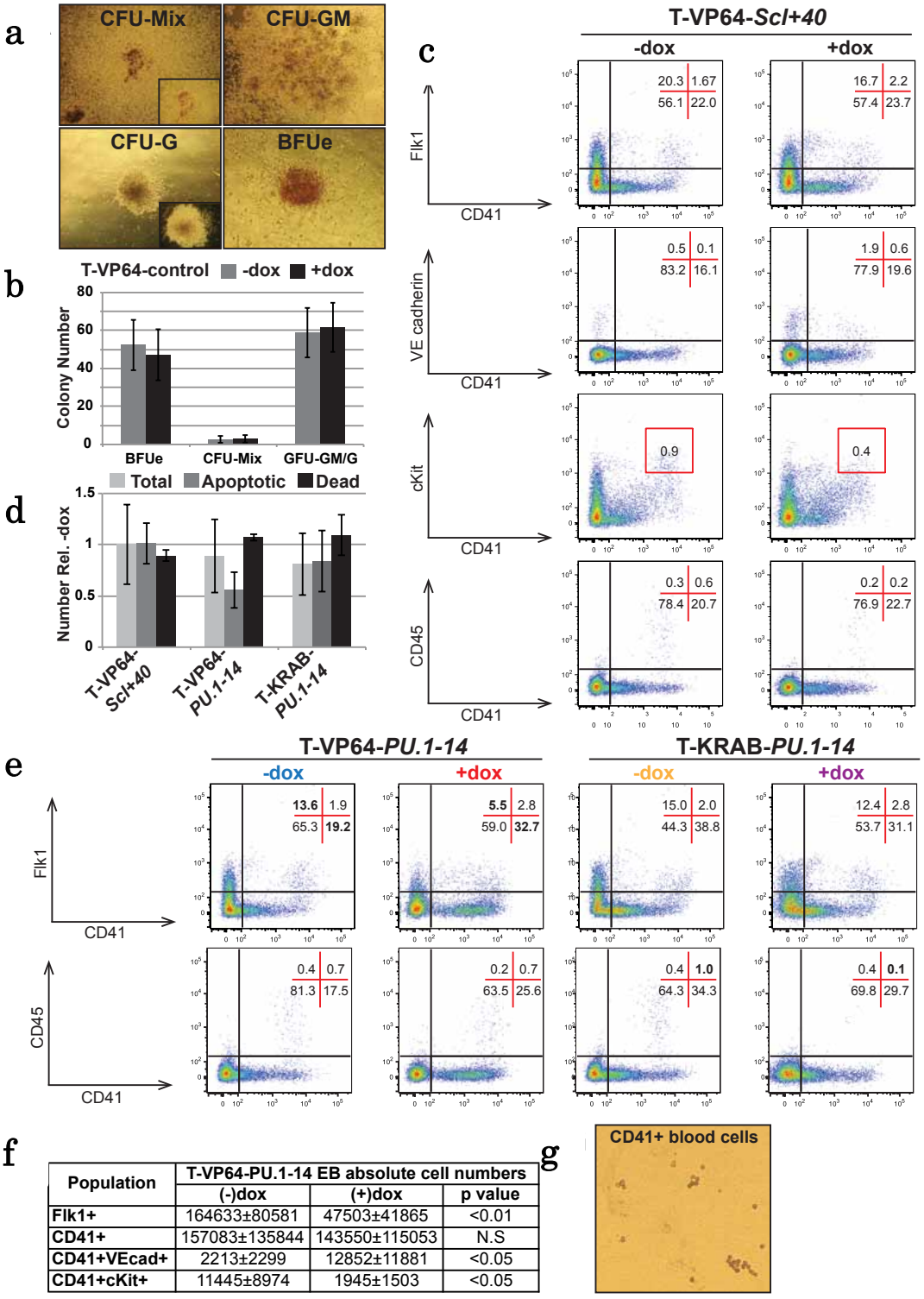
To be confident that TALE-VP64 expression alone was not affecting CFU frequency, ES cell lines carrying a non-functional TALE-VP64 (due to missense mutations in its DNA binding domain) were assessed, which were generated previously

(Gao et al., 2013). Non-functional TALE-VP64 expression did not affect CFU frequency in this assay (**Figure 3.7b**), and confirmed phenotypic changes observed by TALE expression are due to TALE-mediated gene expression perturbations. To correlate the changes in haematopoietic progenitors/CFUs with changes in the cellular composition of the day 6 EBs, day 6 EBs were analysed by flow cytometry. Consistent with the modest effects in colony forming assays, expression of the TALE-VP64 targeting the *Scl+40kb* element minimally affected haematopoietic cell populations present in day 6 EBs (**Figure 3.7c**). Expression of TALEs targeting the *PU.1-14kb* marginally, but not significantly reduced total cell numbers recovered (**Figure 3.7d**). However, this was not due to increased apoptosis as assessed by Annexin V and DAPI staining of day 6 EBs (**Figure 3.7d**).

Although TALE-VP64 mediated upregulation of *PU.1* caused an increase in the relative size of the $CD41^+$ population (**Figure 3.6c** and 3.**7e**), when combined with total cell numbers recovered from the EBs, this did not result in a significant increase in absolute $CD41^+$ cells (**Figure 3.7f**). Interestingly, TALE-VP64 mediated *PU.1* expression caused a loss of the $Flk1^+$ (mesoderm) population (**Figure 3.7e, f**), and significantly increased the $CD41^+$VE $cadherin^+$ (committing haematopoietic cells) population (by over 5-fold in absolute cell numbers; **Figure 3.6c** and **3.7f**). Additionally, TALE-VP64 mediated *PU.1* upregulation caused a loss of the $CD41^+cKit^{hi}$ ("early definitive haematopoietic progenitor") population, that may help explain the loss of colony forming potential described above (**Figure 6b, c**). Combined with the CFU

assays, these data suggested that *PU.1* upregulation may push differentiating cells towards a haematopoietic fate but then inhibits proliferation of the resulting blood cells. Consistent with this hypothesis, TALE-mediated PU.1 induction modestly increased the numbers of day 4 EB derived colonies containing budding $CD41^+$ haematopoietic cells by 1.5-fold (**Figure 3.6d** and **3.7g**), while PU.1 repression modestly reduced their frequency.

In contrast, the major change caused by downregulation of *PU.1* by the TALE-KRAB targeting the *PU.1-14kb* enhancer was an almost complete loss of the $CD45^+$ population (committed definitive haematopoietic cells; **Figure 3.6a**). The above results caused us to speculate that the delayed haematopoiesis caused by *PU.1* repression might be masking an increase in haematopoietic CFU frequency. To test this further, day 6 EB cells were allowed to mature on OP9s for 24 hours before assessing CFU frequency (**Figure 3.6g**). This led to a significant 3-fold and 9-fold increase in BFUe and CFU-Mix colonies, respectively, consistent with published data that suggest PU.1 expression restricts haematopoietic cells to a myeloid fate (Mak et al., 2011). Combined, these data suggest upregulation of PU.1 drives haematopoietic commitment, but causes loss of proliferative ability within the haematopoietic population, while temporary downregulation of PU.1 inhibits the maturation and differentiation of early haematopoietic cells.



Population	T-VP64-PU.1-14 EB absolute cell numbers		
	(-)dox	(+)dox	p value
Flk1+	164633±80581	47503±41865	<0.01
CD41+	157083±135844	143550±115053	N.S
CD41+VEcad+	2213±2299	12852±11881	<0.05
CD41+cKit+	11445±8974	1945±1503	<0.05

Figure 3.7 | Transient TALE Expression Affects Haematopoietic Cell Fate Decision.

(a) *Representative images of haematopoietic colonies scored in methylcellulose CFU assays in Figures 6b and 7b.*

(b) *Representative haematopoietic colonies numbers from 1 x 10^5 day 6 EB cells derived from mouse ES cells inducible expressing a non-functional TALE-VP64 (due to mutations within the DNA binding domain), previously generated (Gao et al., 2013). Dox added to EBs day 4 to induce TALE-VP64 expression. Colonies grown in methylcellulose supplemented with SCF, IL-3, IL-6 and Epo. Error bars are standard deviation of technical triplicates. No statistically significant changes in CFU numbers were seen from three biological triplicates, as determined by the student t test.*

(c) *Flow cytometry plots of day 6 EB cells showing (from top to bottom) Flk1 vs. CD41, VE cadherin (VEcad) vs. CD41, cKit vs. CD41, CD45 vs. CD41). Representative staining patterns for a T-VP64-Scl+40 expressing mouse ES cell line, both uninduced (-dox) and induced (from day 4; +dox). Distribution of cells within quadrant/gates shown as percentages.*

(d) *Total number (light grey bars), frequency of apoptotic (Annexin V^+ DAPI$^-$; dark grey bars) and frequency of dead (Annexin V^+ $DAPI^+$; black bars) T-VP64-Scl+40, T-VP64-PU.1-14 and T-KRAB-PU.1-14 expressing EB day 6 cells (+dox from day 4) relative to –dox controls. Error bars are standard deviation of three biological replicates. No statistically significant changes were seen from three biological triplicates, as determined by the student t test.*

(e) *Flow cytometry plots of day 6 EB cells showing Flk1 vs. CD41 (top) and CD41 vs. CD45 (bottom). Representative staining patterns for TALE-VP64 (left) and TALE-KRAB (right) targeting PU.1-14kb clones, both uninduced (-dox) and induced (from day 4; +dox). Distribution of cells within quadrant/gates shown as percentages.*

(f) *Table displaying absolute cells numbers for cells populations identified by flow cytometry in Figure 2D ± standard deviation from three biological replicates, and p values (using the student t test). N.S; not significant.*

(g) Representative image of day 4 EB Flk1^{+}-derived colony containing haematopoietic (round, budding) CD41^{+} (stained black) cells scored in Figure 6d. Colonies containing endothelial cells that stained weakly CD41+ were not scored unless haematopoietic cells were also present.

3.3 Single Cell Gene Expression Analysis of TALE-mediated PU.1 Perturbation

Having determined the phenotypic effects of TALE-mediated *PU.1* expression perturbations by both colony assays and flow cytometry, next was asked what effects *PU.1* modulation might have on TF regulatory networks. To investigate this, and the phenotypic pro-haematopoietic bias caused by *PU.1* expression in mesoderm, the effect of TALE-VP64 expression was assessed on induction of 44 haematopoietic, mesoderm and endothelial TFs and surface markers as well as four control housekeeping genes in single day 6 EB VE cadherin^{+} (VEcad^{+}) cells using the Fluidigm Biomark platform. At this time point, VEcad expression marks endothelium and haemagenic endothelial precursors, which were not expected to express robust *PU.1* levels. To provide an internal control, a chimeric mixture of the wild type (WT) Ainv18 control and TALE inducible ES cells were differentiated, and sorted VEcad^{+} cells from mCherry and mCherry^{+} populations at day 6 (48 hours after dox addition; **Figure 3.8a**). The expression of all 48 genes in 160 single cells were assessed for each population, which after quality control, resulted in expression data for 136 and 147 cells, respectively, corresponding to a total of over 13,000 RT-qPCR expression scores.

PU.1 was only expressed in 33% (45 of 136) of mCherry^{-} VEcad^{+} cells (**Figure 3.8b**). By contrast, TALE-VP64 efficiently induced *PU.1* expression to 84%

(124 of 147) of the mCherry^{+} VEcad^{+} cells. Moreover, *PU.1* expressing cells in the mCherry^{+} VEcad^{+} population tended to express *PU.1* at a higher level than the *PU.1* expressing cells in the mCherry^{-} VEcad^{+} population (an average of 3.3 ΔCT higher, relative to *Polr2a* and *Ubc* expression; **Figure 3.8b**). This observation demonstrates that the TALE-VP64 can induce gene expression from the *PU.1-14kb* efficiently (but not with complete efficiency), and that the distribution of *PU.1* expression levels within PU.1 expressing cells is altered with a much larger proportion of individual cells expressing high levels of *PU.1*. Single cell expression analysis therefore reveals both qualitative consequences (shift towards more cells expressing) as well as quantitative consequences (shift towards higher per-cell expression levels) of TALE-mediated activation of *PU.1*.

Importantly, TALE-induced *PU.1* expression was associated with consistent changes in the expression of other genes. For example, mCherry^{+} cells from the differentiated *PU.1-14kb* TALE-VP64 ES cells expressed several other haematopoietic genes at higher levels, such as *Csf1r*, *Gata3*, *Gfi1b*, *Runx1* and *Tel/Etv6* (**Figure 3.8b**), suggesting TALE-VP64 induced *PU.1* expression precociously activates a haematopoietic TF network. Interestingly, mCherry^{+} cells also express less of several genes thought to be important for mesoderm or haemogenic endothelium, including *Ets1*, *Etv2*, *Flk1*, *Notch1*, *Sox17* and *VEcad* (**Figure 3.8b**). Moreover, gene expression changes measured by RT-qPCR for *Flk1*, *CD41*, and *Kit* correlated well with expression of these surface markers by flow cytometry (**Figure 3.6c, 3.8b, 3.7 and 3.9**). As Kit

encodes the receptor for the pro-proliferative cytokine Stem Cell Factor (Scf), its downregulation at the transcriptional level and cell surface may partially explain the loss proliferative ability in *PU.1-14kb* TALE-VP64 induced day 6 EB cells (**Figure 3.6d**).

Pairwise all-against-all comparisons of the expression of the 44 TFs and surface proteins across all 283 single cells were performed by calculating Spearman rank correlation coefficients, which were displayed using a heatmap to illustrate both positive and negative correlations between pairs of genes. This identified two positively correlated gene clusters, a haematopoietic gene cluster (including *PU.1*), and a mesodermal/ endothelial gene cluster (**Figure 3.8c**). Although genes from both clusters can be co-expressed in single cells (**Figure 3.8d**), genes from the haematopoietic cluster predominantly showed negative correlation to genes from the haematopoietic cluster predominantly showed negative correlation to genes from the endothelial cluster (**Figure 3.8c**), suggesting an antagonism between these regulatory networks. Pairwise analysis and hierarchical clustering of cells based on their gene expression signatures largely separated the mCherry and mCherry^{+} cells within the VEcad^{+} population (**Figure 3.8d**). As expected, it was within the mCherry^{+} population that the positively correlated cluster of haematopoietic genes is more frequently activated, while expression of the mesodermal/endothelial gene cluster is downregulated.

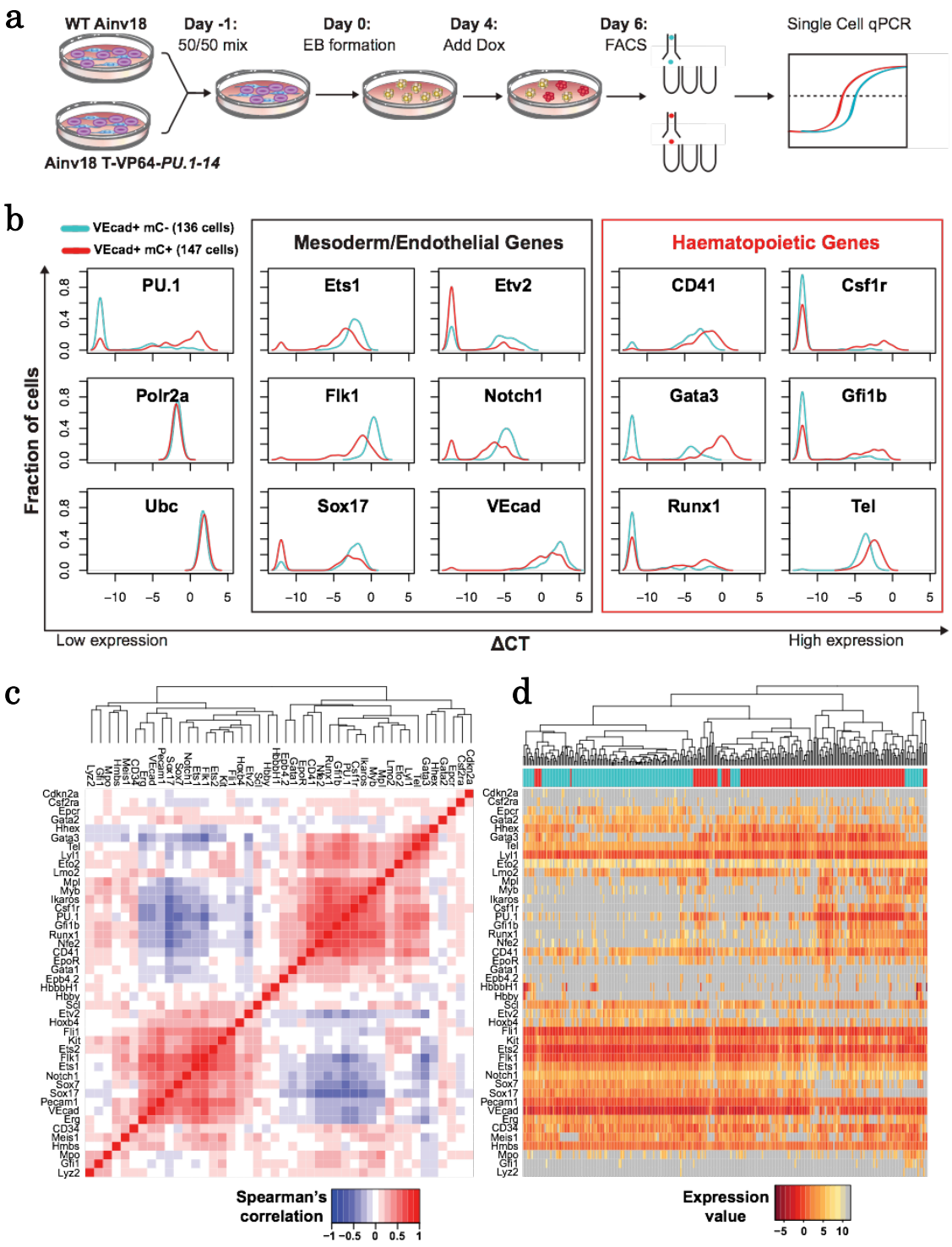


Figure 3.8| Single Cell Analysis of TALE-Mediated PU.1 Expression in Haematopoietic Precursors Suggests a Role in the Transition from an Endothelial to Haematopoietic Transcriptional Programme

(a) *Strategy for single cell gene expression analysis of TALE-mediated perturbations. Wild type (WT) Ainv18 and TALE-VP64 (targeting the PU.1-14kb) targeted ES cells were passaged once as a 50/50 mix before EB formation. Dox was added at day 4 and EBs disaggregated at day 6. Single VE cadherin^{+} (VEcad^{+}) cells (mCherry^{+} and mCherry^{-} sorted as TALE-VP64 expressing and WT, respectively) were sorted into 96 well PCR plates containing lysis buffer, RT/Taq and primers for pre-amplification. Single tube reverse transcription and targeted pre-amplification was undertaken, followed by multiplexed qPCR gene expression analysis using the Fluidigm platform.*

(b) *Density plots of gene expression in day 6 EB VEcad^{+} mCherry^{-} (136 WT Ainv18; in cyan) and VEcad^{+} mCherry^{+} (147 Ainv18 expressing TALE-VP64 targeting PU.1-14kb; in red). The density indicates the fraction of cells at each expression level, relative to housekeeping genes (Polr2a and Ubc). Cells with non-detected gene expression set to -12. See Figure SS for density plots for all 48 genes analysed in these two populations.*

(c) *Hierarchical clustering of Spearman rank correlations between all pairs of genes (excluding housekeepers) from all 283 VEcad^{+} cells (red, positive correlation; blue, negative correlation).*

(d) *Hierarchical clustering of the 283 VEcad^{+} cells according to gene expression (across) with genes ordered according to (C) (Dark red, highly expressed; grey, non-expressed). Top bar indicates cell type: cyan, mCherry-; red, mCherry+.*

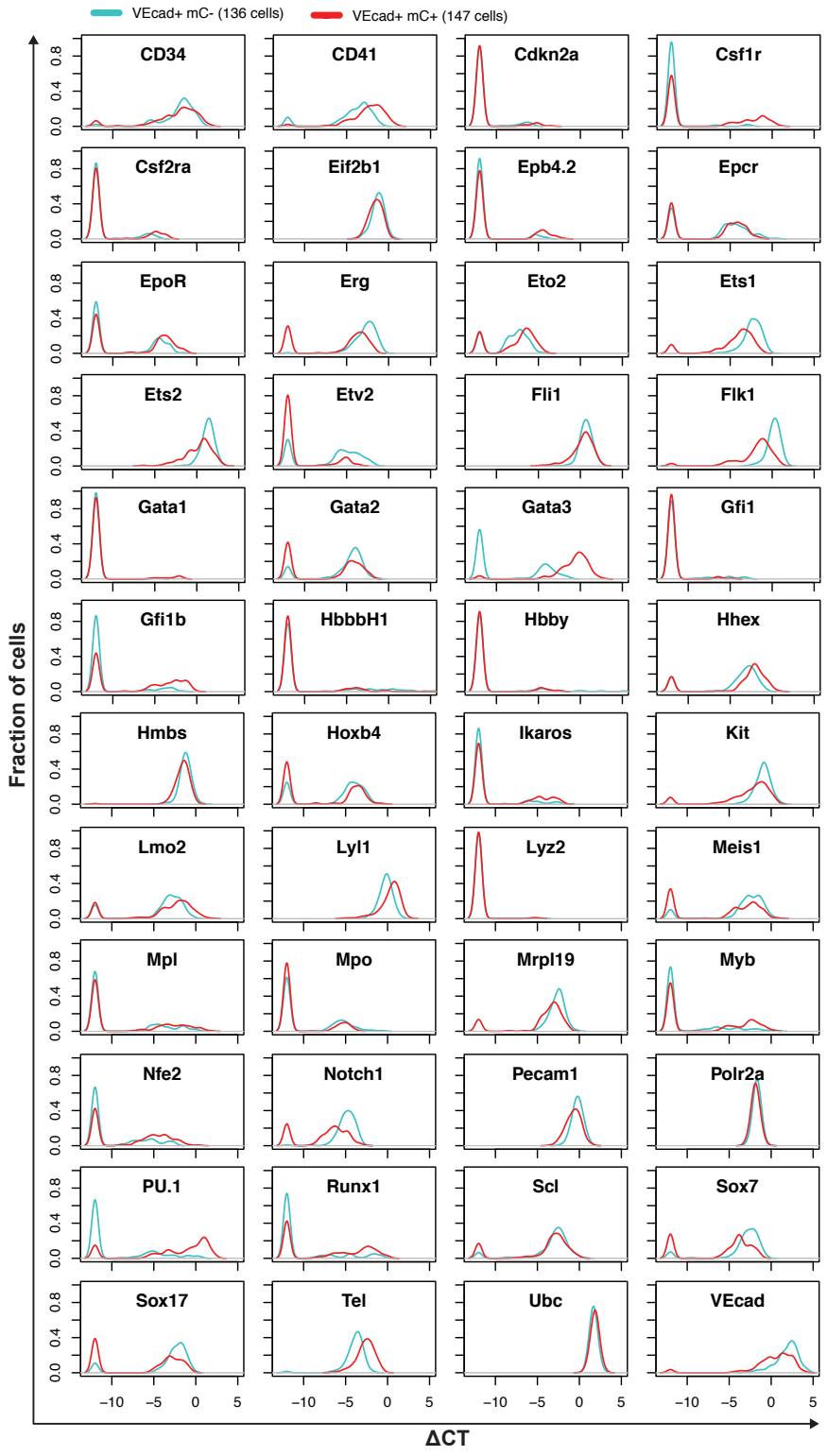


Figure 3.9| Single Cell Analysis of TALE-mediated PU.1 Expression, related to Figure 3.8.

Density plots of gene expression in day 6 EB VEcad$^+$ mCherry$^-$ (136 WT Ainv18; in cyan) and VEcad$^+$ mCherry$^+$ (147 Ainv18 expressing T-VP64-PU.1-14; in red) for all 48 genes analysed. The density indicates the fraction of cells at each expression level, relative to housekeeping genes (Polr2a and Ubc). Cells with non-detected gene expression set to -12.

3.4 *PU.1* can Promote Haematopoietic Commitment of Haemogenic Endothelial Precursors

The data described above suggest that precocious *PU.1* expression in haematopoietic precursors can drive haematopoietic commitment through activation of a haematopoietic TF network. To investigate this further, single cell gene expression analyses for the CD41$^+$cKit^{hi} population was performed additional from both WT Ainv18 and the *PU.1-14kb* TALE-KRAB differentiated ES cells. As above, 160 mCherry$^+$ and mCherry$^-$ cells were sorted at day 6 of culture, from which 142 and 141 single cells, respectively, passed quality control. Within the CD41$^+$cKit^{hi} mCherry$^-$ (WT Ainv18) population, over 90% expressed *PU.1* (132 of 141), and clearly had acquired a committed haematopoietic gene expression pattern (including *Runx1*, *Myb* and *Ikaros*) with only a few cells expressing mesoderm/endothelial-associated genes (*e.g. Sox7*, *Sox17* and *Etv2*; **Figure 3.10a, 3.11**). By contrast, less than 60% (85 of 142) of the CD41$^+$cKit^{hi} mCherry$^+$ (*PU.1-14kb* TALE-KRAB) cells expressed detectable *PU.1* transcript, and *PU.1* was expressed at lower levels in those that did (an average of 2.8 ΔCT lower; **Figure 3.10a**), thus demonstrating that the TALE-KRAB efficiently repressed *PU.1* expression in CD41$^+$cKit^{hi} cells. Expression of *Csf1r*, a known

downstream target of *PU.*1 is tightly correlated with *PU.1* expression, and is not expressed in cells lacking *PU.1* (**Figure 3.10a**). Other genes affected by repression of *PU.1* in $CD41^{+}cKit^{hi}$ included downregulation of *Ikaros* and *Lyl1*, as well as upregulation of *Erg*, *Gata2* and *Myb* or increasing the fraction of cells expressing the respective genes (**Figure 3.10a**).

Having generated a total of 566 single expression profiles from the TALE-VP64 and TALE-KRAB perturbation experiments, next all expression levels were combined to explore the potential of this substantial dataset for the identification of potential regulatory relationships between the 48 genes measured. Pairwise all-against-all comparisons were performed as before by calculating Spearman rank correlation coefficients (**Figure 3.10b**). This analysis placed *PU.1* next to a cluster of haematopoietic genes containing amongst others *Myb*, *Runx1* and *Ikaros*. A second cluster of strongly correlating genes consisted of endothelial genes (*e.g. Sox7*, *VE cadherin* and *Pecam1*). *Gata2* was adjacent to a third and somewhat smaller cluster, consisting of erythroid genes as *Gata1*, *Epb4.2* and globin genes. Of note, *PU.1* showed negative correlation with *Gata2* as expected from the results in **Figure 3.10b**, but not with the core erythroid genes such as *Gata1*. Our analysis therefore suggests that the previously reported cross-antagonism between *Gata1* and *PU.1* (Rekhtman et al., 1999; Zhang et al., 1999; Nerlov et al., 2000; Zhang et al., 2000) may not be operative during the early stages of blood cell specification surveyed in this study. In contrast, a negative

correlation of *PU.1* with many genes within the "endothelial" cluster was observed and suggesting that *PU.1* may antagonize endothelial fate.

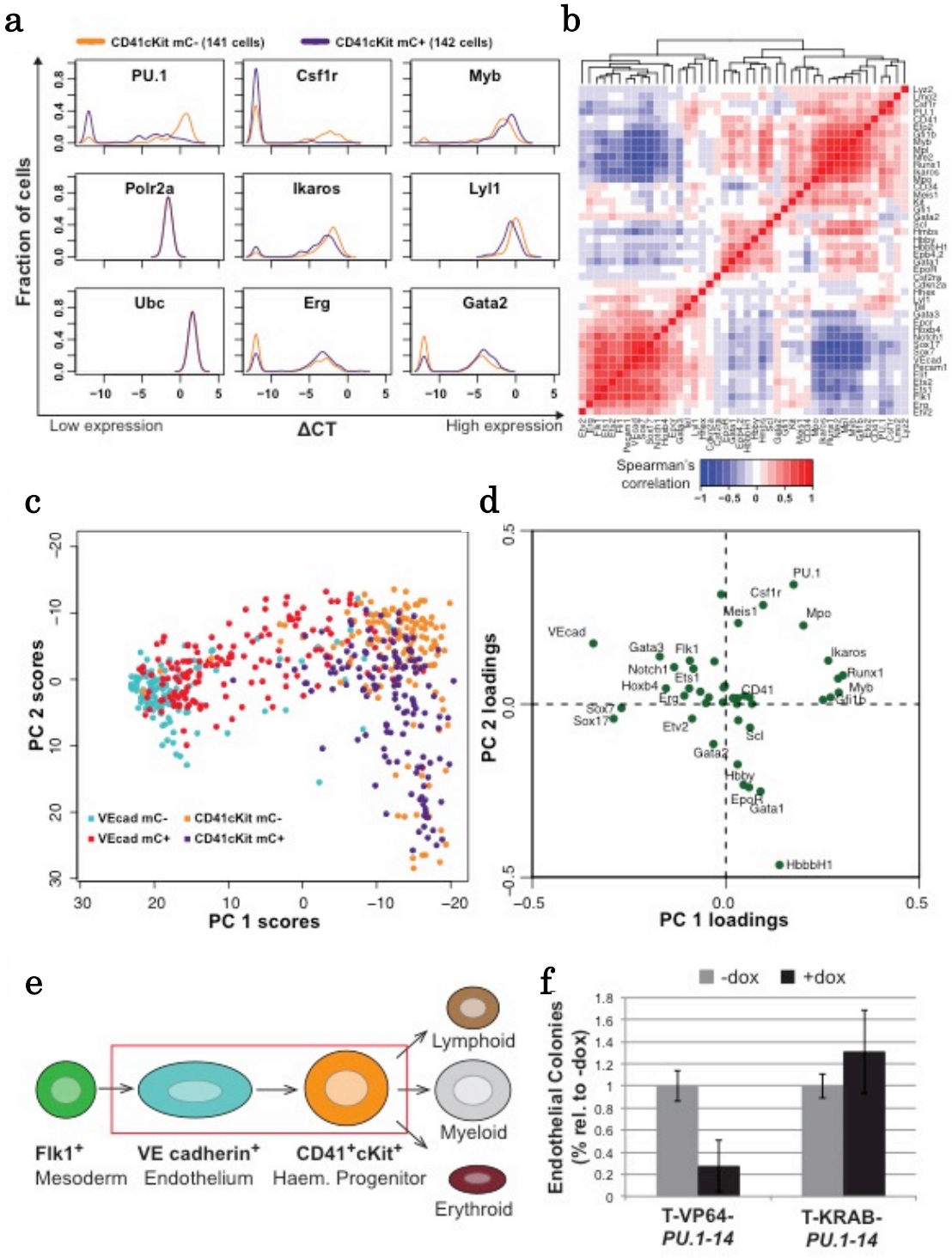


Figure 3.10| TALE-Mediated Expression Perturbations Suggests Transcriptional Interactions During Blood Specification and a Role for PU.1 in Antagonising Endothelial Fate

(a) *Density plots of gene expression in day 6 EB CD41^{+}cKit^{hi} (CD41cKit) mCherry^{-} (141 WT Ainv18; in orange) and CD41cKit mCherry^{+} (142 Ainv18 expressing TALE-KRAB targeting PU.1-14kb; in purple). The density indicates the fraction of cells at each expression level, relative to housekeeping genes (Polr2a and Ubc). Cells with non-detected gene expression set to -12. See* ***Figure 3.11*** *for density plots for all 48 genes analysed in these two populations.*

(b) *Hierarchical clustering of Spearman rank correlations between all pairs of genes (excluding housekeepers) from using gene expression data from all 566 cells (VEcad^{+} and CD41cKit).*

(c) *Principal component analysis (PCA) of the 566 VEcad^{+} and CD41cKit cells, in the first and second components, from the expression of all 44 genes (excluding the four housekeeping genes used as controls).*

(d) *Principal component loadings indicate the extent to which each gene contributes to the separation of cells along each component in (C).*

(e) *Current model of definitive haematopoietic specification from Flk1^{+} mesoderm through a haemogenic endothelial precursor to a haematopoietic stem/progenitor that can differentiate into lymphoid, myeloid or erythroid lineages.*

(f) *Endothelial potential of TALE expressing VEcad^{+} cells, as a percentage of –dox control cells.*

To further assess possible effects of *PU.1* expression perturbations on the entire multi-dimensional gene expression dataset from all 566 cells, Principal Component Analysis (PCA) was preformed. PCA separated VEcad^{+} and CD41^{+}cKit^{hi} mCherry^{-} cells into distinct populations across principle component 1 (PC1), consistent with the notion of two developmentally distinct populations (**Figure 3.10c**), confirmed

also by the PCA loading plot, which showed this separation to be driven by expression of endothelial genes in the VEcad^{+} population (including VEcad, Sox17 and Sox7) and haematopoietic TFs in the CD41^{+}cKit^{hi} population (including *Runx1*, *Myb*, *Gfi1b*, *Ikaros*, and *PU.1*) (**Figure 3.10d**). The CD41^{+}cKit^{hi} population is resolved into two populations by PC2, by expression of myeloid genes (including *PU.1* and *Csf1r*) and erythroid genes (including *Hbb-bH1*, *Gata1* and *EpoR*), thus suggesting the CD41^{+}cKit^{hi} population contains myeloid and erythroid biased CD41^{+}cKit^{hi} progenitor cells. PCA dataset therefore provided good resolution of early developmental populations based on current models of developmental haematopoietic specification (**Figure 3.10e**, based on (**ref.** Medvinsky et al., 2011). Interestingly, TALE-VP64 *PU.1* activated VEcad^{+} mCherry^{+} cells bridge the separation between the control VEcad^{+} and CD41^{+}cKit^{hi} populations (**Figure 3.10c**), consistent with the notion that *PU.1* expression pushes VEcad^{+} cells to haematopoietic commitment, but is unable to drive the transition completely. By contrast, the separation of the TALE-KRAB *PU.1* repressed CD41^{+}cKit^{hi} mCherry^{+} population from the CD41^{+}cKit^{hi} mCherry population is less striking, although more *PU.1* repressed cells are closer to the VEcad^{+} population and none form part of the most distant group of cells in the top right hand part of the plot (**Figure 3.10c**), consistent with the block in haematopoietic maturation observed for these cells.

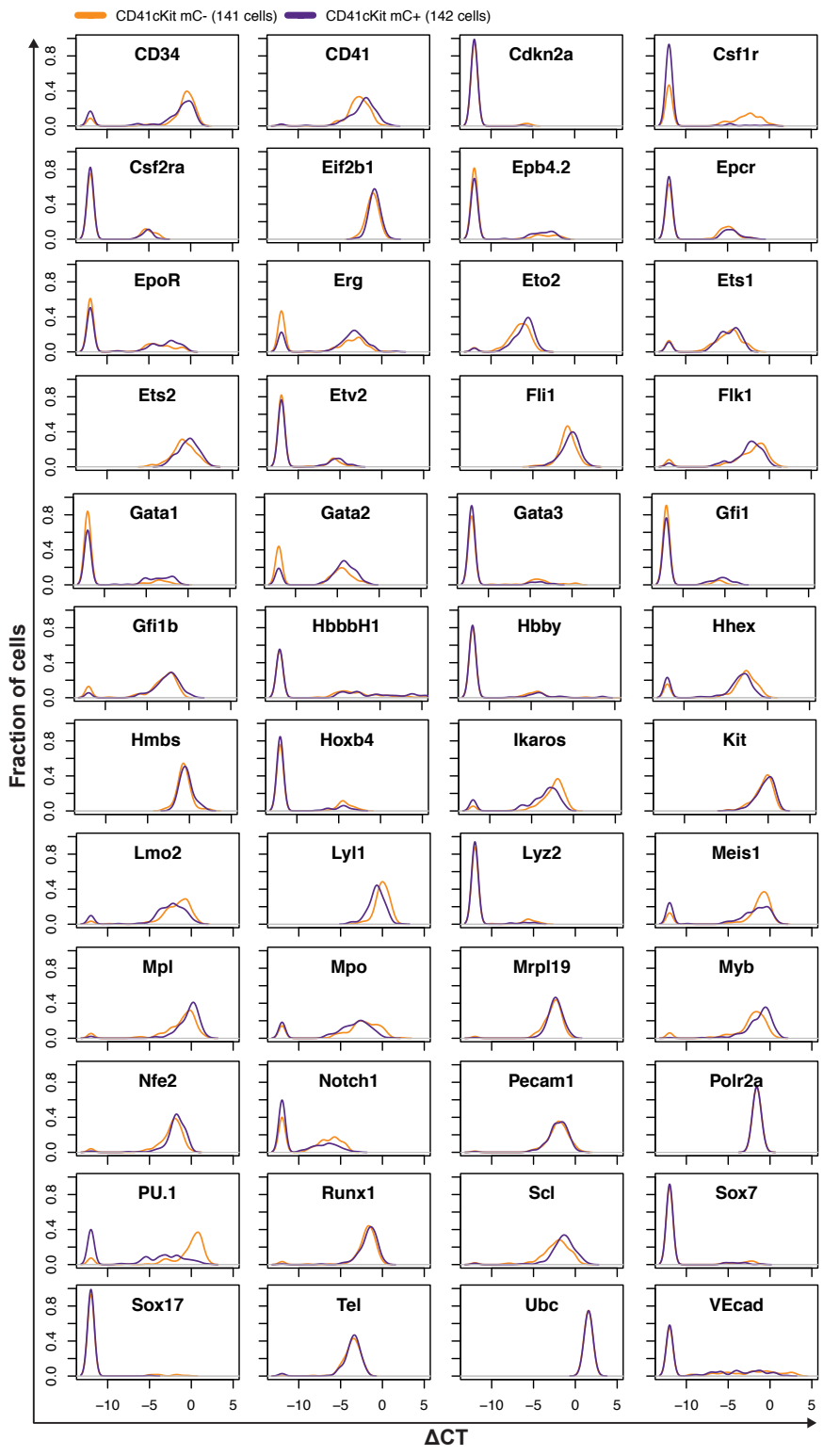


Figure 3.11| related to Figure 3.10

Density plots of gene expression in day 6 EB CD41+cKithi mCherry- (141 WT Ainv18; in orange) and CD41+cKithi mCherry+ (142 Ainv18 expressing T-KRAB-PU.1-14; in purple) for all 48 genes analysed. The density indicates the fraction of cells at each expression level, relative to housekeeping genes (Polr2a and Ubc). Cells with non-detected gene expression set to -12.

Both the pairwise correlation and PCA analyses suggested that *PU.1* expression contributes to a haematopoietic fate in VEcad^{+} cells. Therefore, the effect of *PU.1* perturbation was assessed on endothelial potential of the day 6 VEcad^{+} cells. TALE-VP64 mediated *PU.1* activation inhibited endothelial colony formation, while PU.1 repression did not (**Figure 3.10f** and **3.12**). Combined, these data suggest activation of *PU.1* expression during developmental haematopoiesis plays a role in driving a haematopoietic rather than endothelial transcriptional programme, and activation of *PU.1* expression in haemogenic endothelium may be an important molecular decision in haematopoietic commitment.

Such a large single cell gene expression dataset presented the opportunity to investigate underlying TF network interactions active during the endothelial-to-haematopoietic transition (EHT) using Partial Correlation Analysis. This analysis identifies network interactions (edges) by detecting irreducible statistical dependencies between TFs. To visualise the results, the 34 network edges were plotted between the TF nodes with highly significant correlations (p values <0.0001; **Figure 3.13**). Although this method of analysis provides positive/negative correlation information, directionality cannot be inferred. Most TF interactions were positive and

formed a highly interconnected network, which could be important in network stabilisation. Two types of negative correlations were observed: (1) between haematopoietic-specific and endothelial-specific genes, including *Runx1* and *Sox17*, and (2) between haematopoietic-lineage specific genes, including *Nfe2* and Gata3. Such TF antagonisms may be important switches in cell fate commitment. Interestingly, Runx1 has previously been found to bind at the *Sox17* loci in similar Runx1-expressing haemogenic endothelium (Lichtinger et al., 2012), suggesting such network interaction may be direct. As expected by above results, *PU.1* is a highly connected node positively correlating with haematopoietic genes, consistent with a role for PU.1 in stabilising haematopoietic cell fate decision.

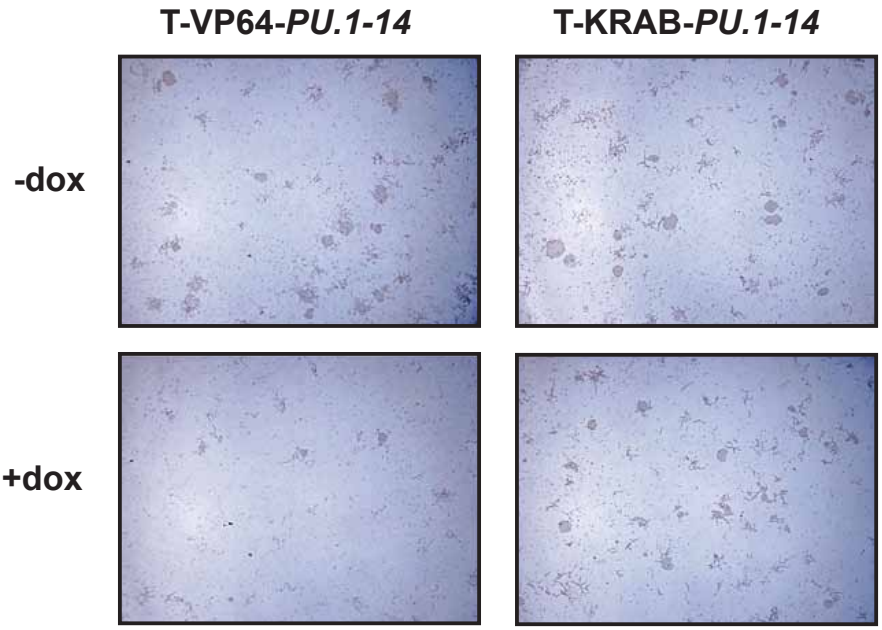


Figure 3.12| related to Figure 3.10

Representative images of endothelial colony assays from day 6 EB VEcad^{+} cells from T-VP64-PU.1-14 or T-KRAB-PU.1-14 ES cell lines as described in Figure 3.13.

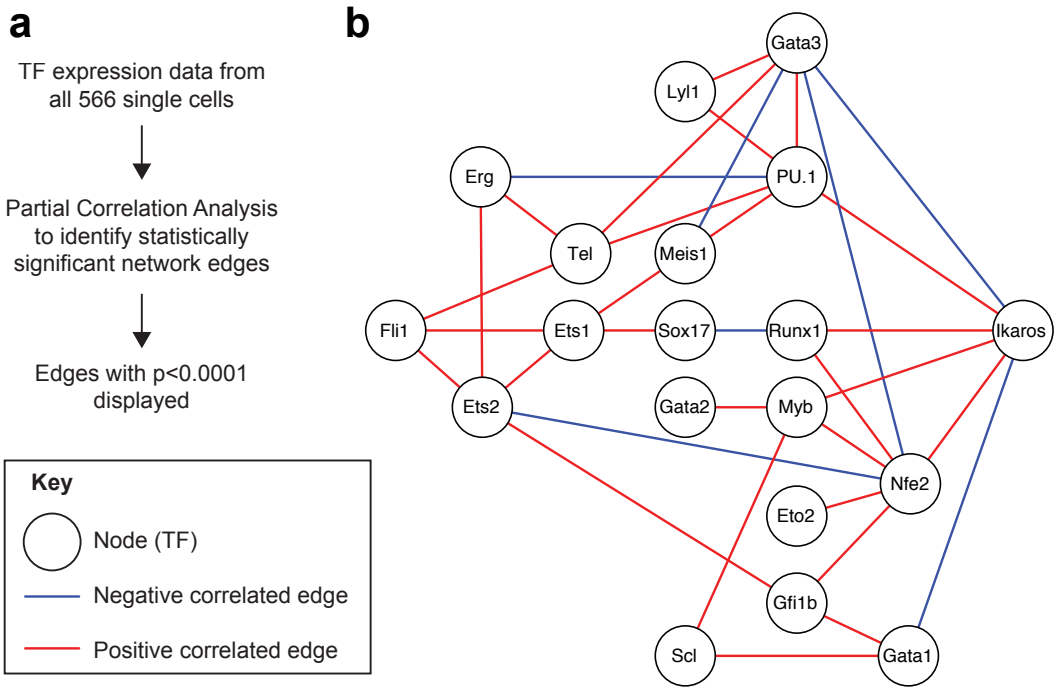


Figure 3.13| Partial Correlation Analysis identified a highly interconnected TF network active during the endothelial-to-haematopoietic transition (EHT)

(a) *Above, schematic of method used to build TF network model in (b) and below, diagram key for (b).* ***(b)*** *TF network model built from Partial Correlation Analysis using Spearman correlation with highly statistically significant interactions (p<0.0001) displayed as connections (edges) between TFs (nodes).*

4 DISCUSSION

Results presented here demonstrate a novel use of TALEs in combination with single cell gene expression profiling to investigate the consequences of transcriptional perturbation on developmental regulatory networks. High-throughput RT-qPCR coupled with comprehensive cellular assays uncovered a previously unrecognised role

for *PU.1* during the development of early haematopoietic progenitors from haemogenic endothelial cells during ES cell differentiation.

The analysis of several hundred single cells with TALE induction highlights the efficiency of TALE-mediated perturbation on endogenous gene expression at the single cell level, which proved to be comparable to alternative methods of perturbation such as siRNA knockdown (Kouno et al., 2013), and provided more physiologically relevant expression changes when comparing upregulation of gene expression using TALE-VP64 proteins with retroviral cDNA overexpression. Moreover, TALE-mediated perturbation does not require distinction between exogenous and endogenous cDNAs, allow normal co-and post-transcriptional processing to occur, and allows for detection by gene expression assays that are located in untranslated regions (UTR).

The CRISPR-Cas9 system has recently been adapted to modulate gene expression by a similar mechanism to TALEs (Gilbert et al., 2013; Maeder et al., 2013; Perez-Pinera et al., 2013; Qi et al., 2013). As the CRISPR-Cas9 target specificity is based on guide RNAs rather than a modular protein domain, generation of these "designer" TFs is faster than assembly of TALEs (Gaj et al., 2013). However, a recent comparison between CRISPR-Cas9 and TALEs suggested higher targeting specificity for the latter (Fu et al., 2013; Mali et al., 2013).

The majority of previous research on PU.1 has been on its role in adult haematopoiesis, where high PU.1 levels promote terminal myeloid differentiation while reduced PU.1 expression results in proliferation, reviewed elsewhere (Mak et al., 2011),

and consistent with CFU data. A novel link between PU.1 levels and the cell cycle has been described recently where PU.1, by regulating cell cycle lengthening, determines PU.1 protein accumulation within the cell, effecting lympho-myeloid cell fate decisions (Kueh et al., 2013). The data additionally highlight the importance of tightly regulated *PU.1* expression for early haematopoiesis to occur. Indeed, the increased haematopoietic progenitor frequency seen by transient repression of *PU.1* may provide a useful model to study cell seft-renewal and differentiation decisions. Moreover, since the TALEs used here target conserved DNA sequences within *cis*-regulatory elements, these tools can be directly applied to manipulate human haematopoiesis, with the additional key advantage that TALE-mediated perturbation can be temporally controlled.

PU.1 has recently been shown to inhibit proliferation by directly controlling cell cycle regulators (Staber et al., 2013), which is consistent with observed loss of haematopoietic colonies after TALE-VP64-mediated *PU.1* upregualtion and increase in haematopoietic colonies after TALE-KRAB-mediated *PU.1* expression. Since the gene selection for single cell expression analysis was focused on TF networks controlling early haematopoietic development, genes other than those assayed are likely to contribute to the phenotypic changes caused by *PU.1* expression perturbations, and this may well include cell cycle regulators. However, this data already suggest novel regulatory relationships, such as a tight positive correlation between *PU.1* and *Lyl1*, and antagonism between *PU.1* and *Sox17* expression. Moreover, while gene expression

correlations can be extracted from WT gene expression data alone, TALE-mediated perturbations provide evidence that such correlations are due to TF network interactions. Interestingly, *PU.1* has recently been suggested to positively regulate *Lyl1* also in foetal thymocytes (Del Real and Rothenberg, 2013). It is worth highlighting that TALE-mediated perturbations caused consistent gene expression changes in single cells suggesting *PU.1* operates within a tightly interconnected haematopoietic TF network.

In adult haematopoiesis, *PU.1-14kb* is a known a target of Runx1 (Huang et al., 2008), a critical TF for definitive haematopoiesis (Okuda et al., 1996; Wang et al., 1996). Wilkinson et al. report that the *PU.1-14kb* element is active *in vivo* in midgestation AGM blood clusters, where definitive HSPCs arise. Using ES cell differentiation assays, Runx1 has been shown to initiate chromatin unfolding at the *PU.1-14kb* early during haematopoietic specification, priming it for later activity (Hoogenkamp et al., 2009). Such enhancer priming is likely to be important for efficient TALE-VP64 mediated induction of expression. While Hoogenkamp et al. were unable to determine the frequency of such priming events within the precursor population, the data presented here would be consistent with a model where the majority of VEcad^{+} cells contain primed *PU.1-14kb* enhancers, due to the high efficiency of TALE-mediated *PU.1* expression activation in this cell type. Enhancer priming may also contribute to the low level expression of *PU.1* prior to haematopoietic commitment, with more robust expression seen later as haematopoietic TF network circuitry is reinforced. Such low level expression may be analogous to transcriptional noise of

lineage regulators previously seen in adult haematopoietic progenitor cells (Pina et al., 2012).

When considered the translational relevance to the dental science fields, this study provided the usefulness of TALEs as a tool for analyzing the relationship between cellular function/differentiation and the target gene expressions, as well as the role of PU.1 in the transcriptional network of cell differentiation. As stated in the introduction, Spi1/PU.1 is important in Cfs1r signaling for driving macrophage differentiation, that is a key cell lineage of osteoimmunology (Lorenzo et al., 2011). In the inflamed periodontal tissue, abundant of immune cells were infiltrated and evoked a variety of immune reactions. It has been well-demonstrated that macrophages were present and activated in the periodontally-affected lesions to play some regulatory roles in osteoclastgenesis by producing inflammatory cytokines (Rayyan, 2013).

A variety of regenerative therapies were clinically introduced in Periodontics, such as bone grafts, guided tissue regeneration treatment, and application of enamel matrix derivatives and signaling molecules for the last quarter-century. However these therapies have the issue in clinical practice, including technique sensitivity, limitations of indications, and the predictability and longevity of outcome (Illueca et al., 2006). Since emerging concept of tissue engineering in 1990s, application of stem cell therapy was also applied experimentally and clinically as the strategies for periodontal regeneration (Sarita et al., 2012). Unfortunately, this newly-developed therapy still has the issues and was not enabled to achieve the complete regeneration. To date, at least

five different dental stem cells; dental pulp stem cells, stem cells from exfoliated deciduous tissues, periodontal ligament stem cells, stem cells from apical papilla, and dental follicle cells, were applied for tissue engineering in the craniofacial area other than bone marrow and adipose tissue-derived stem cells. However, there are a number of technical issues for using these cells as source of cell-therapy, including isolation of appropriate cell types, establishment of optimal growth and differentiation, and the delivery for transplantationRef4). Also the craniofacial regeneration including tooth and periodontal tissue bioengineering has a unique characteristic to regenerate many types of hard and soft tissues with a 3D-configuration. In this point of view, Eleuterio et al. (2013) recently reported that stem cells from periodontal ligament and dental pulp differentially expressed proteins as compared to bone marrow stem cells, suggesting the different cell lineage using proteome analysis. Thus, it is very important to find the key signaling molecules and realize the detail of signaling pathway regulating oral tissue development. TALEs may be provide as a useful tool for revealing the full scope of these signaling pathways.

In summary, I have validated use of TALEs targeting conserved *cis*-regulatory elements as an efficient, multifaceted tool to modulate endogenous gene expression and study TF regulatory networks perturbations in single cells, and in doing so have uncovered a role for PU.1 in haematopoietic specification. Understanding the physiological relevance and significance of myeloid cells development under steady-state condition regulated by PU.1 may facilitate the modification of treatment strategies

for periodontal wound healing in a diabetic inflammatory environment. Further studies are needed to elucidate the underlying molecular mechanisms and interactions during myeloid cells development, including the roles between Runx1 and Sox17 and Nfe2 and Gata3, other regulatory factors, and additional characteristics that endow myeloid subsets with their pro-inflammatory and hypersensitive potential such as their lineage (*e.g.* Diabetic-derived macrophages) and origin.

5 REFERENCES

Alpiste-Illueca FM, Buitrago-Vera P, de Grado-Cabanilles P, Fuenma-yor-Fernandez V, Gil-Loscos FJ. (2006) .Periodontal regeneration in clinical practice. *Med Oral Patol Oral Cir Bucal*;11:E382-92.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology* **215**, 403-410.

Beerli, R. R., Segal, D. J., Dreier, B. and Barbas, C. F., 3rd. (1998). Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proc Natl Acad Sci U S A* **95**, 14628-14633.

Boch, J. and Bonas, U. (2010). Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annual review of phytopathology* **48**, 419-436.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A. and Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326**, 1509-1512.

Bogdanove, A. J. and Voytas, D. F. (2011). TAL effectors: customizable proteins for DNA targeting. *Science* **333**, 1843-1846.

Cong, L., Zhou, R., Kuo, Y. C., Cunniff, M. and Zhang, F. (2012). Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nature communications* **3**, 968.

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., Hanna, J., Lodato, M. A., Frampton, G. M., Sharp, P. A. et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**, 21931-21936.

Dakic, A., Wu, L. and Nutt, S. L. (2007). Is PU.1 a dosage-sensitive regulator of haemopoietic lineage commitment and leukaemogenesis? *Trends in immunology* **28**, 108-114.

Del Real, M. M. and Rothenberg, E. V. (2013). Architecture of a lymphomyeloid developmental switch controlled by PU.1, Notch and Gata3. *Development* **140**, 1207-1219.

Dexter, T. M., Allen, T. D., Scott, D. and Teich, N. M. (1979). Isolation and characterisation of a bipotential haematopoietic cell line. *Nature* **277**, 471-474.

Enrica Eleuterio , Oriana Trubiani , Marilisa Sulpizio, Fabrizio Di Giuseppe, Laura Pierdomenico, Marco Marchisio, Raffaella Giancola, Gianluigi Giammaria, Sebastiano Miscia, Sergio Caputi, Carmine Di Ilio, Stefania Angelucci (2013). Proteome of Human Stem Cells from Periodontal Ligament and Dental Pulp. *PlosOne* DOI: 10.1371/journal.pone.0071101

Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K. and Sander, J. D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature biotechnology* **31**, 822-826.

Gaj, T., Gersbach, C. A. and Barbas, C. F., 3rd. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in biotechnology* **31**, 397-405.

Gao, X., Yang, J., Tsang, J. C., Ooi, J., Wu, D. and Liu, P. (2013). Reprogramming to Pluripotency Using Designer TALE Transcription Factors Targeting Enhancers. *Stem cell reports* **1**, 183-197.

Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E., Stern-Ginossar, N., Brandman, O., Whitehead, E. H., Doudna, J. A. et al. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442-451.

Graf, T. and T. Enver. (2009). Forcing cells to change lineages. *Nature.* **462**:587-594.

Guo, G., Huss, M., Tong, G. Q., Wang, C., Li Sun, L., Clarke, N. D. and Robson, P. (2010). Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Developmental cell* **18**, 675-685.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H. and Glass, C. K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell* **38**, 576-589.

Hoogenkamp, M., Lichtinger, M., Krysinska, H., Lancrin, C., Clarke, D., Williamson, A., Mazzarella, L., Ingram, R., Jorgensen, H., Fisher, A. et al. (2009). Early chromatin unfolding by RUNX1: a molecular explanation for differential requirements during specification versus maintenance of the hematopoietic gene expression program. *Blood* **114**, 299-309.

Huang, G., Zhang, P., Hirai, H., Elf, S., Yan, X., Chen, Z., Koschmieder, S., Okuno, Y., Dayaram, T., Growney, J. D. et al. (2008). PU.1 is a major downstream target of AML1 (RUNX1) in adult mouse hematopoiesis. *Nature genetics* **40**, 51-60.

Keller, G., Kennedy, M., Papayannopoulou, T. and Wiles, M. V. (1993). Hematopoietic commitment during embryonic stem cell differentiation in culture. *Mol Cell Biol* **13**, 473-486.

Kent, W. J. (2002). BLAT--the BLAST-like alignment tool. *Genome Res* **12**, 656-664.

Knezevic, K., Bee, T., Wilson, N. K., Janes, M. E., Kinston, S., Polderdijk, S., Kolb-Kokocinski, A., Ottersbach, K., Pencovich, N., Groner, Y. et al. (2011). A Runx1-Smad6 rheostat controls Runx1 activity during embryonic hematopoiesis. *Mol Cell Biol* **31**, 2817-2826.

Kouno, T., de Hoon, M., Mar, J. C., Tomaru, Y., Kawano, M., Carninci, P., Suzuki, H., Hayashizaki, Y. and Shin, J. W. (2013). Temporal dynamics and transcriptional control using single-cell gene expression analysis. *Genome biology* **14**, R118.

Kueh, H. Y., Champhekhar, A., Nutt, S. L., Elowitz, M. B. and Rothenberg, E. V. (2013). Positive feedback between PU.1 and the cell cycle controls myeloid differentiation. *Science* **341**, 670-673.

Kyba, M., Perlingeiro, R. C. and Daley, G. Q. (2002). HoxB4 confers definitive lymphoid-myeloid engraftment potential on embryonic stem cell and yolk sac hematopoietic progenitors. *Cell* **109**, 29-37.

Lichtinger, M., Ingram, R., Hannah, R., Muller, D., Clarke, D., Assi, S. A., Lie, A. L. M., Noailles, L., Vijayabaskar, M. S., Wu, M. et al. (2012). RUNX1 reshapes the epigenetic landscape at the onset of haematopoiesis. *The EMBO journal* **31**, 4318-4333.

Lorenzo, J., Horowitz, Choi Y., Schett, G., Takanayanagi, H. (2011) Osteoimmunology: Interactions of the Immune and Skeletal Systems. **Academic Press**. ISBN: 978-0123756701.

Lozzio, B. B., Lozzio, C. B., Bamberger, E. G. and Feliu, A. S. (1981). A multipotential leukemia cell line (K-562) of human origin. *Proc Soc Exp Biol Med* **166**, 546-550.

Maeder, M. L., Linder, S. J., Cascio, V. M., Fu, Y., Ho, Q. H. and Joung, J. K. (2013). CRISPR RNA-guided activation of endogenous human genes. *Nature methods* **10**(10):977-9.

Mak, K. S., Funnell, A. P., Pearson, R. C. and Crossley, M. (2011). PU.1 and Haematopoietic Cell Fate: Dosage Matters. *International journal of cell biology* **2011**, 808524.

Mali, P., Aach, J., Stranges, P. B., Esvelt, K. M., Moosburner, M., Kosuri, S., Yang, L. and Church, G. M. (2013). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature biotechnology* **31**, 833-838.

Moignard, V., Macaulay, I. C., Swiers, G., Buettner, F., Schutte, J., Calero-Nieto, F. J., Kinston, S., Joshi, A., Hannah, R., Theis, F. J. et al. (2013). Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis. *Nature cell biology* **15**, 363-372.

Nakano, T., Kodama, H. and Honjo, T. (1994). Generation of lymphohematopoietic cells from embryonic stem cells in culture. *Science* **265**, 1098-1101.

Nerlov, C., Querfurth, E., Kulessa, H. and Graf, T. (2000). GATA-1 interacts with the myeloid PU.1 transcription factor and represses PU.1-dependent transcription. *Blood* **95**, 2543-2551.

Okuda, T., van Deursen, J., Hiebert, S. W., Grosveld, G. and Downing, J. R. (1996). AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. *Cell* **84**, 321-330.

Okuno, Y., Huang, G., Rosenbauer, F., Evans, E. K., Radomska, H. S., Iwasaki, H., Akashi, K., Moreau-Gachelin, F., Li, Y., Zhang, P. et al. (2005). Potential autoregulation of transcription factor PU.1 by an upstream regulatory element. *Mol Cell Biol* **25**, 2832-2845.

Papantonis, A. and Cook, P. R. (2013). Transcription factories: genome organization and gene regulation. *Chemical reviews* **113**, 8683-8705.

Pimanda, J. E. and Gottgens, B. (2010). Gene regulatory networks governing haematopoietic stem cell development and identity. *International Journal of Developmental Biology* **54**, 1201-1211.

Perez-Pinera, P., Kocak, D. D., Vockley, C. M., Adler, A. F., Kabadi, A. M., Polstein, L. R., Thakore, P. I., Glass, K. A., Ousterout, D. G., Leong, K. W. et al. (2013). RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nature methods* **10**(10):973-6.

Pina, C., Fugazza, C., Tipping, A. J., Brown, J., Soneji, S., Teles, J., Peterson, C. and Enver, T. (2012). Inferring rules of lineage commitment in haematopoiesis. *Nature cell biology* **14**, 287-294.

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P. and Lim, W. A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173-1183.

Rayyan A. Kayal. (2013). The Role of Osteoimmunology in Periodontal Disease. *BioMed Research International*, v. 2013, Article ID 639368.

Rosenbauer, F., Wagner, K., Kutok, J. L., Iwasaki, H., Le Beau, M. M., Okuno, Y., Akashi, K., Fiering, S. and Tenen, D. G. (2004). Acute myeloid leukemia induced by graded reduction of a lineage-specific transcription factor, PU.1. *Nature genetics* **36**, 624-630.

Sarita Dabra, Kamalpreet Chhina, Nitin Soni, and Rakhi Bhatnagar (2012). Tissue engineering in periodontal regeneration: A brief review. *Dent Res J (Isfahan)* 9(6): 671–680.

Schutte, J., Moignard, V. and Gottgens, B. (2012). Establishing the stem cell state: insights from regulatory network analysis of blood stem cell development. *Wires Syst Biol Med* **4**, 285-295.

Sroczynska, P., Lancrin, C., Pearson, S., Kouskoff, V., Lacaud, G. (2009). In vitro differentiation of mouse embryonic stem cells as a model of early hematopoietic development. *Methods Mol Biol.* **538**:317-34.

Staber, P. B., Zhang, P., Ye, M., Welner, R. S., Nombela-Arrieta, C., Bach, C., Kerenyi, M., Bartholdy, B. A., Zhang, H., Alberich-Jorda, M. et al. (2013). Sustained PU.1 levels balance cell-cycle regulators to prevent exhaustion of adult hematopoietic stem cells. *Molecular Cell* **49**, 934-946.

Toriello, N. M., Douglas, E. S., Thaitrong, N., Hsiao, S. C., Francis, M. B., Bertozzi, C. R. and Mathies, R. A. (2008). Integrated microfluidic bioprocessor for single-cell gene expression analysis. *Proc Natl Acad Sci U S A* **105**, 20173-20178.

Wang, Q., Stacy, T., Binder, M., Marin-Padilla, M., Sharpe, A. H. and Speck, N. A. (1996). Disruption of the Cbfa2 gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive hematopoiesis. *Proc Natl Acad Sci U S A* **93**, 3444-3449.

Wang, W., Lin, C., Lu, D., Ning, Z., Cox, T., Melvin, D., Wang, X., Bradley, A. and Liu, P. (2008). Chromosomal transposition of PiggyBac in mouse embryonic stem cells. *Proc Natl Acad Sci U S A* **105**, 9290-9295.

Wilkinson, A. C. and Gottgens, B. (2013). Transcriptional regulation of haematopoietic stem cells. *Advances in experimental medicine and biology* **786**, 187-212.

Wilkinson, A. C., Goode, D. K., Cheng, Y. H., Dickel, D. E., Foster, S., Sendall, T., Tijssen, M. R., Sanchez, M. J., Pennacchio, L. A., Kirkpatrick, A. M. et al. (2013). Single site-specific integration targeting coupled with embryonic stem cell differentiation provides a high-throughput alternative to in vivo enhancer analyses. *Biology open* **2**, 1229-1238.

Zhang, F., Cong, L., Lodato, S., Kosuri, S., Church, G. M. and Arlotta, P. (2011). Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nature biotechnology* **29**, 149-153.

Zhang, P., Behre, G., Pan, J., Iwama, A., Wara-Aswapati, N., Radomska, H. S., Auron, P. E., Tenen, D. G. and Sun, Z. (1999). Negative cross-talk between hematopoietic regulators: GATA proteins repress PU.1. *Proc Natl Acad Sci U S A* **96**, 8705-8710.

Zhang, P., Zhang, X., Iwama, A., Yu, C., Smith, K. A., Mueller, B. U., Narravula, S., Torbett, B. E., Orkin, S. H. and Tenen, D. G. (2000). PU.1 inhibits GATA-1 function and erythroid differentiation by blocking GATA-1 DNA binding. *Blood* **96**, 2641-2648.